

DV Assignment 2: Scientific and Information Visualizations

Prateek Rath

IMT2022017

IIIT Bangalore

prateek.rath@iiitb.ac.in

Mohit Naik

IMT2022076

IIIT Bangalore

mohit.naik@iiitb.ac.in

Anurag Ramaswamy

IMT2022103

IIIT Bangalore

anurag.ramaswamy@iiitb.ac.in

INTRODUCTION

Through this assignment we have tried to implement scientific visualization methods, such as color maps, contour plots and quiver plots, as well as information visualization methods, such as node-link diagrams, treemaps and parallel coordinate plots, to understand two different datasets and derive insights.

SCIENTIFIC VISUALIZATIONS

The dataset used is the GRIDMET dataset [1], which contains high-resolution meteorological data, including temperature, humidity, precipitation, wind speed etc. for the United States over a period of 40-45 years. We chose a 3-month window from June to August 2021 for our analysis, because the 2021 U.S. summer was marked by a series of significant and extreme meteorological events across multiple regions - heatwaves, record-breaking temperatures, drought conditions, and tropical storms, all of which offer an interesting ground for visualizations and analysis [2]. There was no pre-processing involved - most of the variables in the datasets had no null values, so we just downloaded the datasets for the year 2021, and clipped them to the required time window.

A. Scalar Field Visualization using Color Maps

Between June 1st and August 31st, 2021, the United States experienced a series of severe weather events, including intense heatwaves, wildfires, and hurricanes. Using color maps, we aimed to visually capture temperature patterns and other weather indicators over this period, allowing us to identify and better understand these events. This section explains the various color map techniques we experimented with and how they helped in making inferences about the weather events that occurred during this period.

1) *Experimenting with Color Map Techniques:* We chose TMMX (Maximum Near-Surface Air Temperature) as the variable on which we performed the experiments as it serves as a primary variable and offers an intuitive representation of daily temperature extremes, making it an effective choice for exploring and interpreting weather patterns. For visualizing this data, various experiments were conducted to determine the most effective mapping techniques. Our primary goal was to ensure that the visualizations accurately represented temperature variations, highlighted critical temperature extremes, and remained easily interpretable. The experiments were divided

into three main categories: color palette selection, scaling approaches, and discretization methods.

1) *Color Palette Selection:* We began by experimenting with various color palettes using temperature data from June 28th, 2021 - a date chosen specifically for its extreme temperature contrasts during the Pacific Northwest heat wave. This notable weather event provided an ideal test case for evaluating how different color palettes represent temperature variations, and we focused on three main categories of palettes, each offering distinct advantages for temperature visualization. All color palettes mentioned below used a consistent scale across all maps, with the global minimum (-10°C) and maximum (55°C) temperature values derived from the entire three-month period of June-August 2021.

- Sequential Palettes such as viridis (Fig.1), YlOrRd (Fig.2), and Reds (Fig.3) were chosen for their smooth, gradual changes in color intensity. These palettes are well-suited for temperature data that increases in a single direction (e.g., low to high temperatures). They are also perceptually uniform, meaning that the change in color reflects the change in temperature consistently across the entire range, making it easier to spot differences.
- Diverging Palettes including RdBu_r (Fig.6), RdYIBu_r (Fig.5), and seismic (Fig.4) were tested to highlight deviations from a central, neutral value (such as an average temperature). These palettes use contrasting colors (e.g., blue for cooler areas and red for hotter areas) to emphasize extreme hot and cold temperatures. This makes them especially effective for identifying unusual temperature events.
- Perceptually Uniform Palettes like magma (Fig.9), inferno (Fig.8), and cividis (Fig.7) were explored for their consistent visual differences across the entire temperature range. These palettes are designed to minimize visual distortion and are particularly beneficial when working with datasets that have a wide range of temperatures. They help avoid misleading color interpretations, but they may lack the familiar warm-to-cool transition that people often associate with temperature data.

After evaluating these options, we chose the seismic color palette due to its strong contrast, which is well-suited for highlighting the extreme characteristic of the June-August 2021 heatwave. The intuitive red-blue gradient helps to clearly distinguish hot and cold areas, making it easier to identify and analyze temperature anomalies during this period.

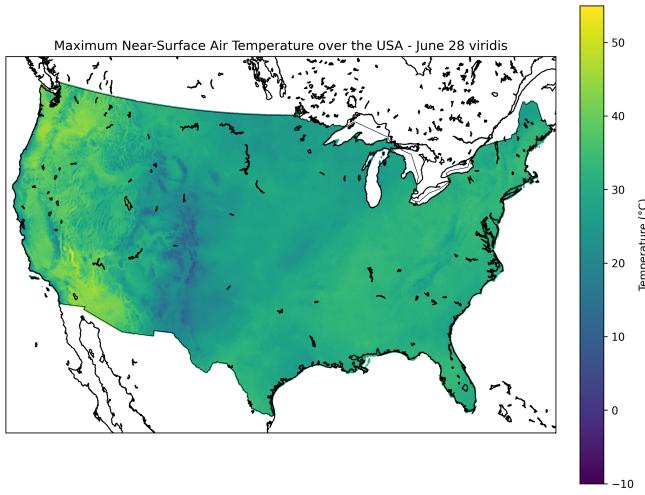


Fig. 1. June 28th - viridis

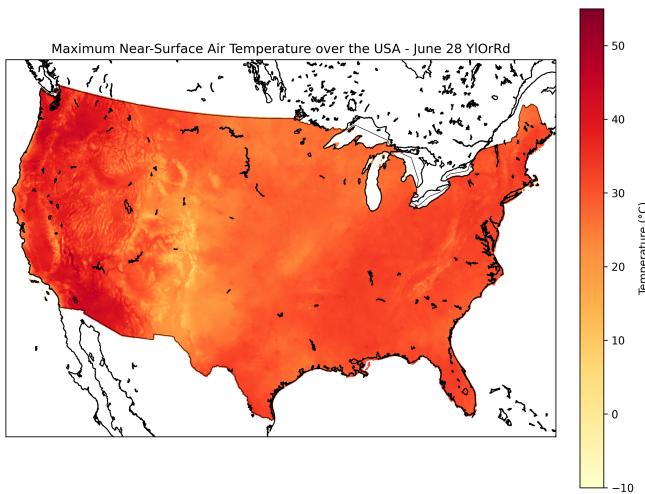


Fig. 2. June 28th - YlOrRd

- 2) Scaling Approaches: After selecting 'seismic' as our color palette, we explored different scaling methods to understand how they affect the visualization of temperature variations. Each scaling approach offers a distinct perspective on the same underlying data, highlighting different aspects of the temperature distribution. Each scaling approach is seen on 3 different days - June 2nd, July 16th and August 31st from Figures 10 to 18.

- Global Scaling uses the minimum and maximum temperatures from the entire three-month period as

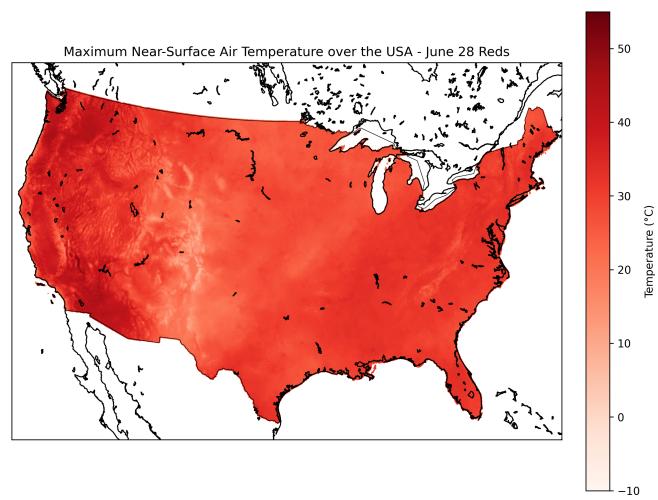


Fig. 3. June 28th - Reds

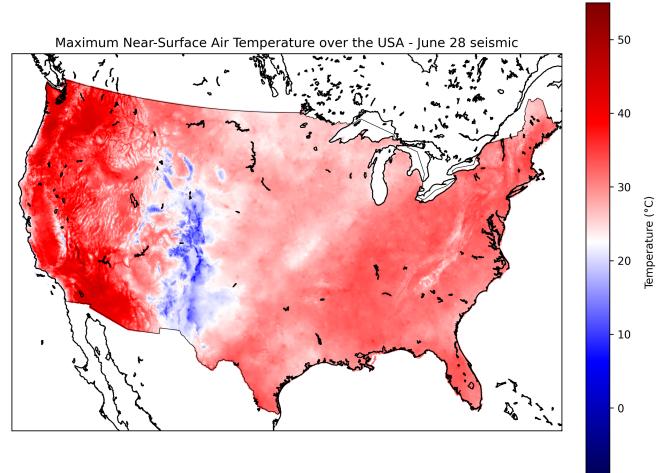


Fig. 4. June 28th - seismic

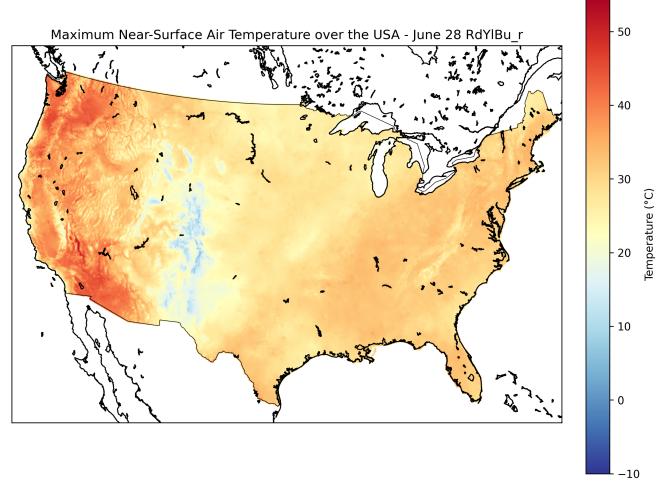


Fig. 5. June 28th - RdYlBu_r

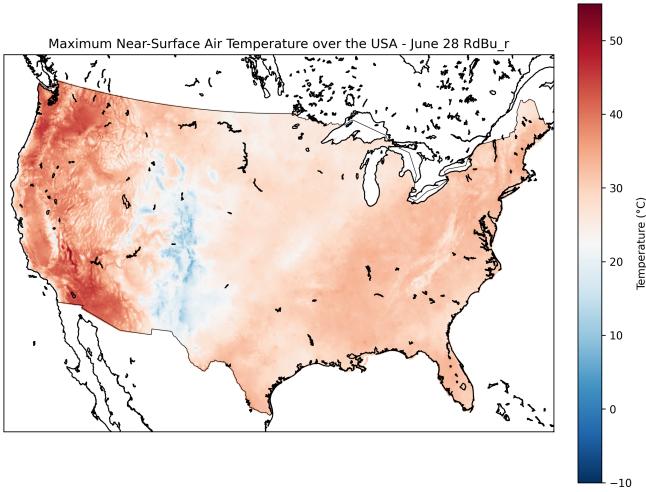


Fig. 6. June 28th - RdBu_r

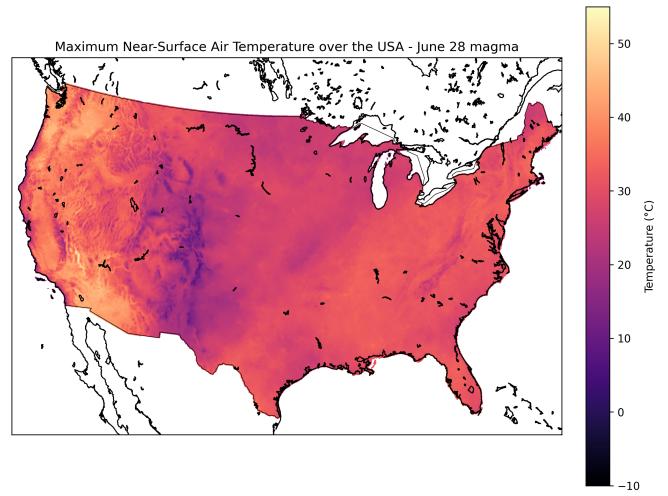


Fig. 9. June 28th - magma

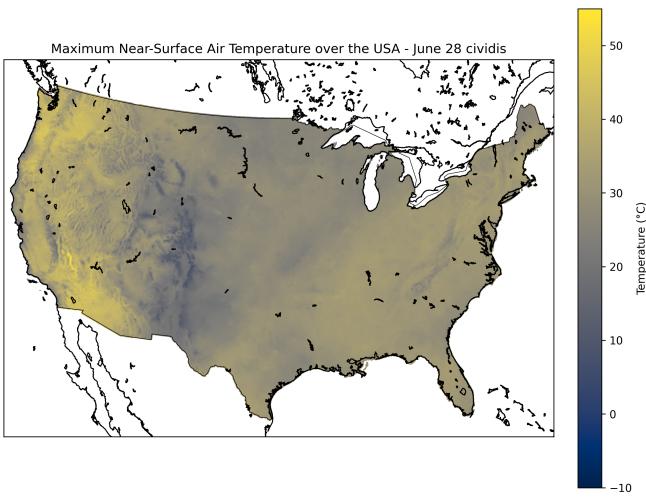


Fig. 7. June 28th - cividis

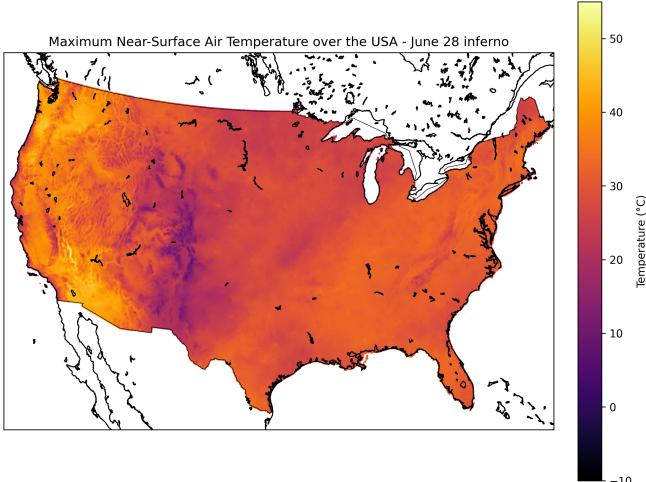


Fig. 8. June 28th - inferno

fixed endpoints for the color scale. This approach maintains consistency across all visualizations, making it easier to compare temperatures between different days and regions. While this method excels at preserving the absolute temperature relationships, it may sacrifice local detail in regions with less extreme variations.

- Local Scaling adjusts the color scale endpoints to match the minimum and maximum temperatures present in each individual day's data. This method maximizes the use of the available color range for each visualization, making it easier to discern subtle temperature variations within a single day. However, it makes direct comparisons between different days more challenging since the same color might represent different temperatures across different visualizations.
- Logarithmic Scaling involves applying a logarithmic transformation to the temperature data before mapping it to colors. While this method can be effective for visualizing data with exponential trends, temperature does not exhibit such a relationship. Consequently, logarithmic scaling was less impactful for our dataset, as it did not enhance the interpretability of temperature variations.

For our analysis of the heat wave, we primarily relied on global scaling as it provided the most accurate representation of the temperature extremes while maintaining comparability across different days.

- 3) Discretization Methods: We experimented with different ways of representing the continuous temperature data by grouping it into distinct categories. Discretization can make it easier to spot specific temperature ranges, but it also risks losing some of the finer details of the data.

- Continuous colormaps as seen in Figure 4 keeps the temperature data in its purest form, mapping every

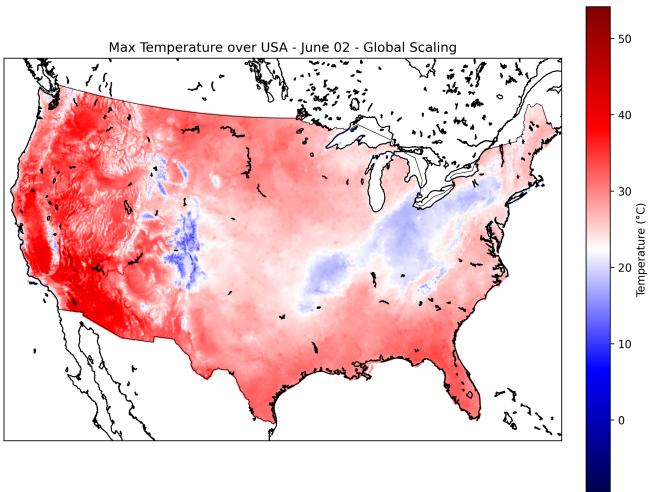


Fig. 10. June 2nd - Global scaling

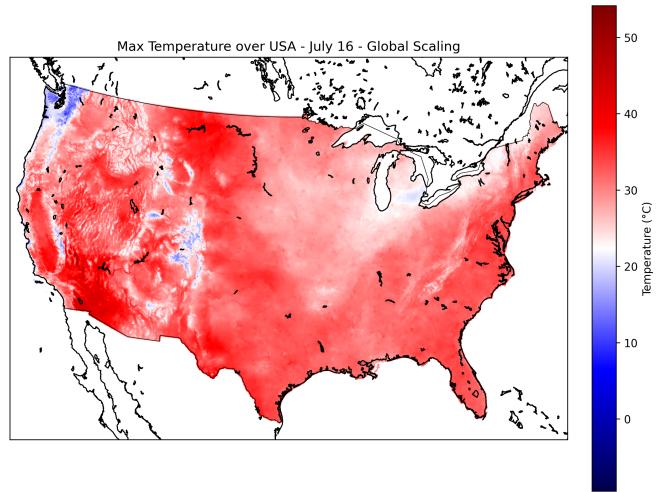


Fig. 13. July 16th - Global scaling

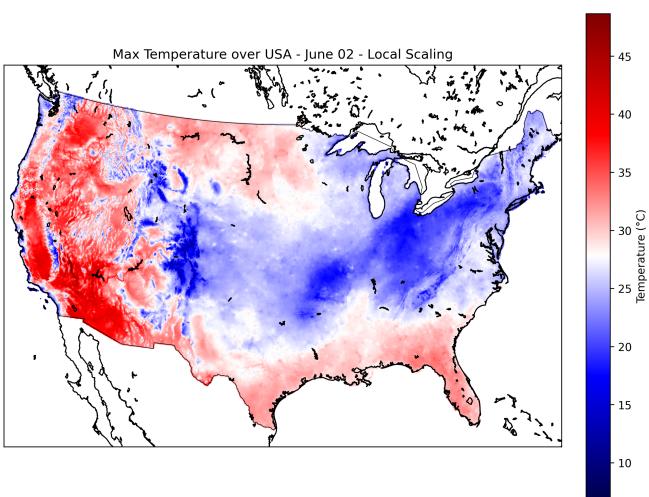


Fig. 11. June 2nd - Local scaling

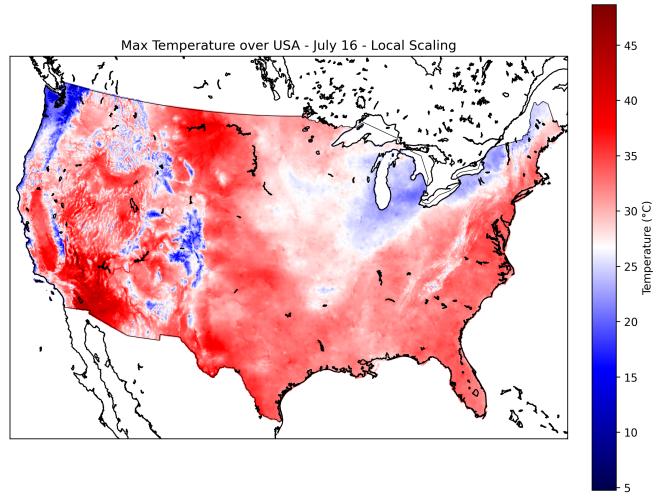


Fig. 14. July 16th - Local scaling

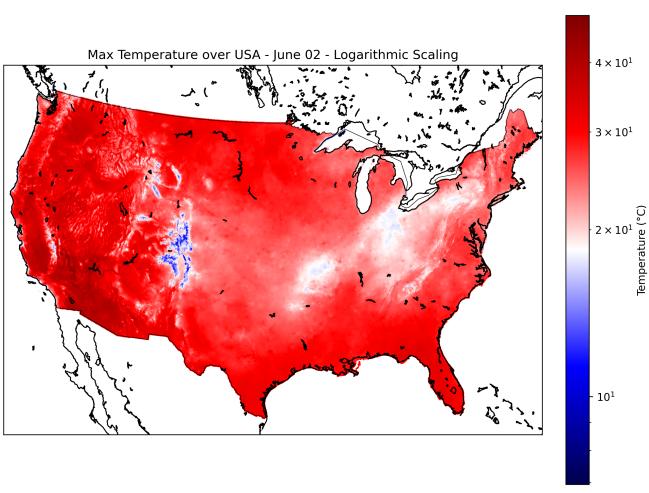


Fig. 12. June 2nd - Logarithmic scaling

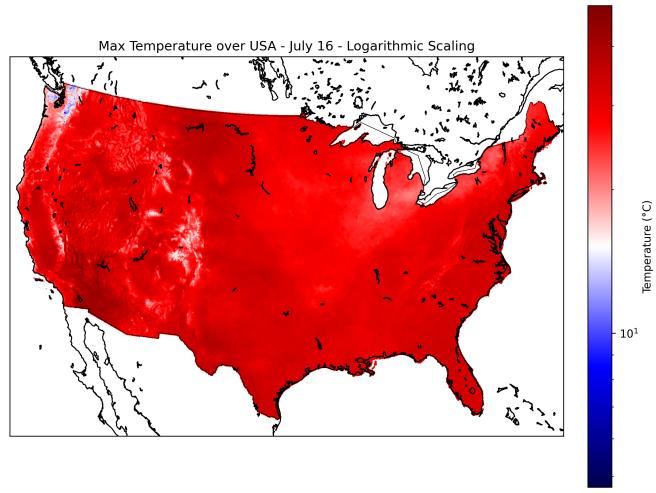


Fig. 15. July 16th - Logarithmic scaling

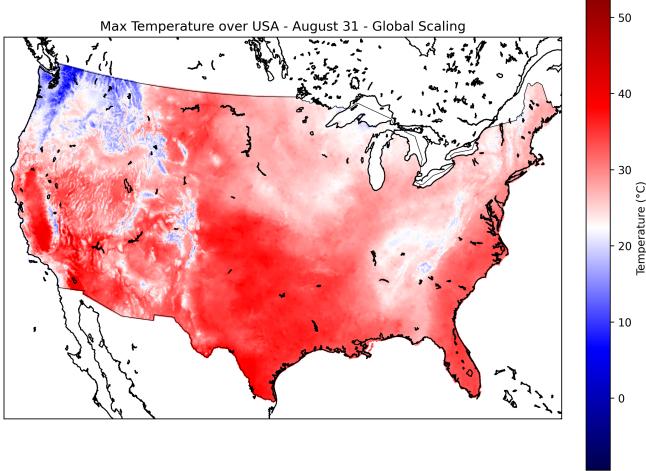


Fig. 16. August 31st - Global scaling

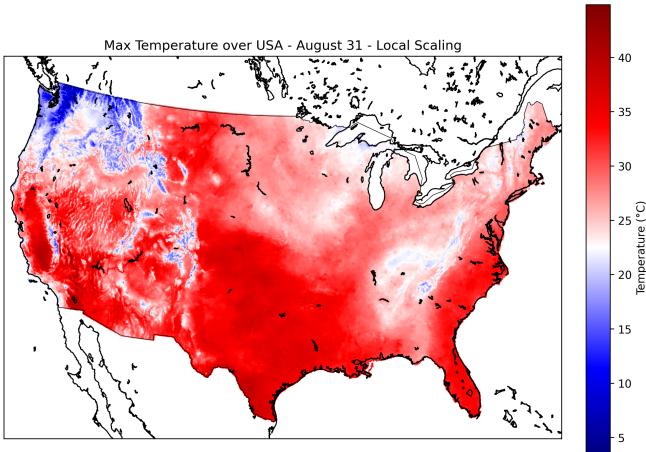


Fig. 17. August 31st - Local scaling

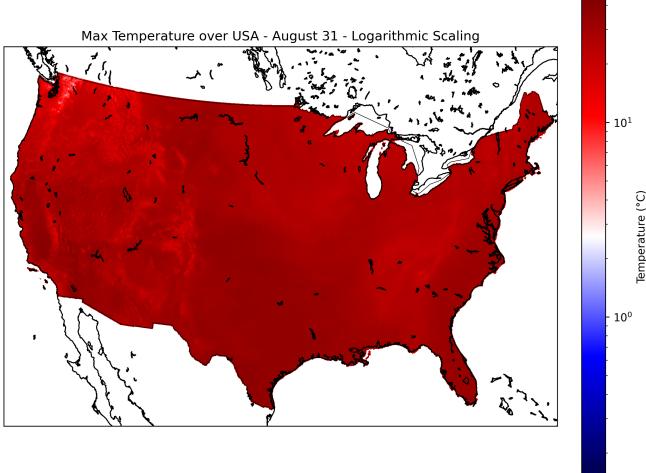


Fig. 18. August 31st - Logarithmic scaling

unique value directly to a color. This provides the most detailed and nuanced view, revealing even the smallest temperature shifts. While it can be a bit tricky to read precise values without a color bar for reference, this method ensures no information is lost, which is particularly important when analyzing complex events like heatwaves.

- 5-Degree Intervals as seen in Figure 19 break the data into broader chunks, grouping temperatures into bands of 5°C each. This method offers a good compromise, making it easier to recognize and communicate key temperature ranges without oversimplifying. However, it may smooth over some of the subtle variations that could be critical in extreme weather analysis.
- 10-Degree Intervals as seen in Figure 20 simplify the visualization even further by creating wider temperature bands. This approach is great for quickly spotting overall patterns and trends but tends to obscure more detailed fluctuations, making it less suitable for analyzing rapid changes during intense events like heatwaves.

For our analysis, we chose the continuous representation because it allowed us to capture the full complexity of the temperature variations during June-August 2021. This choice gave us a clearer picture of the localized spikes in temperature, helping to better understand the dynamics of the heatwave event.

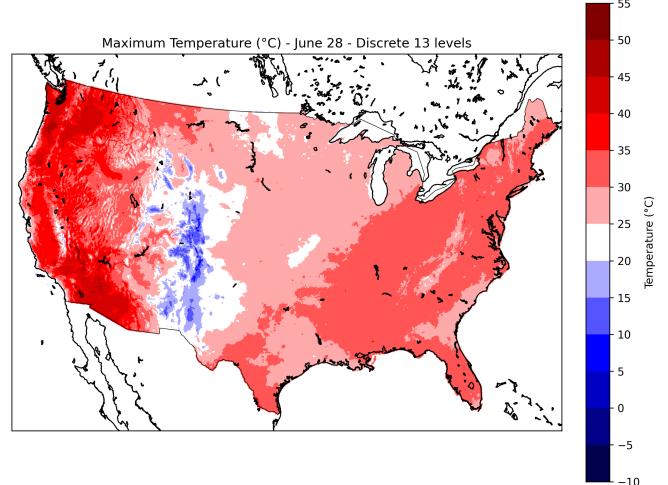


Fig. 19. June 28th - Discretization using 5 degree intervals

2) Analysis of Significant Weather Events Using Visualization Techniques: In this section, we analyze several significant weather events that occurred between June and August 2021 in the United States, using colormaps. In our analysis, we focused on several key variables that play a critical role in understanding weather conditions. TMMX (maximum near-surface air temperature), which tracks the highest temperatures at the Earth's surface, was visualized using a seismic palette

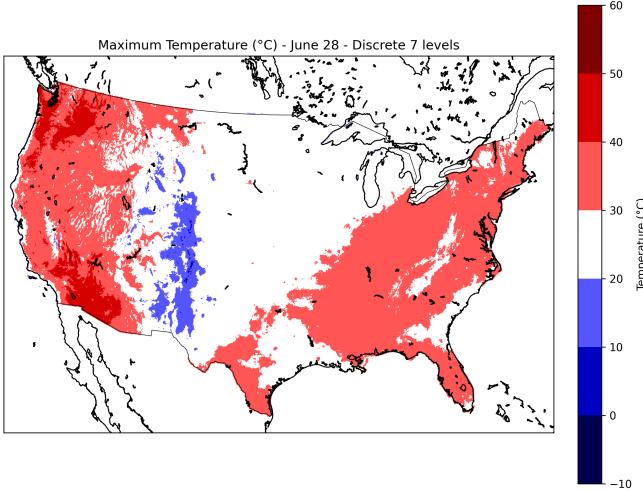


Fig. 20. June 28th - Discretization using 10 degree intervals

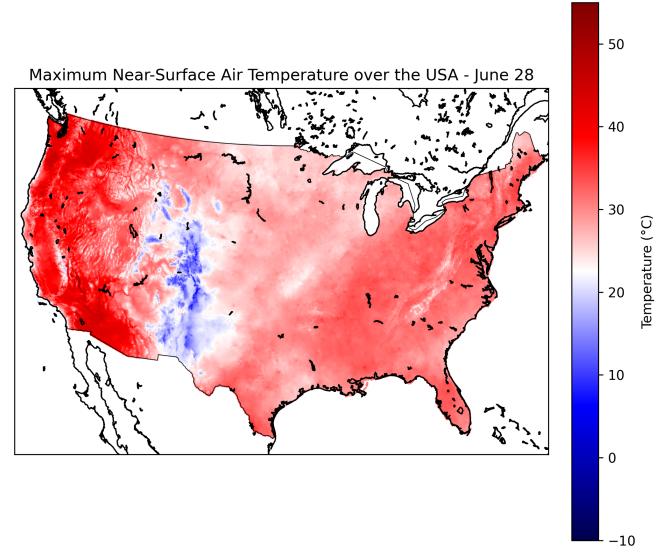


Fig. 21. June 28th - Maximum temperature map using TMMX data

to clearly highlight temperature extremes through a striking blue-to-red gradient. SRAD (surface downwelling shortwave radiation), representing the amount of solar energy hitting the surface, was mapped with the YlOrRd palette to effectively show rising solar intensity. ERC (Energy Release Component) and BI (Burning Index), both indicators of fire danger, were also visualized with YlOrRd, making it easy to see how fire risk escalated during the heatwave. Finally, FM1000 (1000-hour Dead Fuel Moisture), which measures long-term moisture levels in fuel, was displayed using the BrBG diverging palette to highlight the critical shift between wet and dry conditions in fuels.

a) *Pacific Northwest Heat Wave (June 25-30, 2021; data snapshot from June 28, 2021)* [3]: The Pacific Northwest experienced an unprecedented heat wave in late June 2021, with record-breaking temperatures affecting Oregon and Washington. Using our seismic color palette with global scaling, we visualized the maximum near-surface air temperature (TMMX) distribution as seen in Figure 21. The visualization revealed intense red coloring over western Washington and Oregon, indicating temperatures exceeding 45°C in some areas. Complementing the temperature visualization, we employed the YlOrRd palette to map the surface downwelling shortwave radiation (SRAD) as seen in Figure 22. This revealed significantly higher radiation levels across the Columbia River Basin and Willamette Valley, helping explain the mechanism behind the heat wave. The correlation between high SRAD values and extreme temperatures supports the understanding of how trapped solar radiation contributed to this historic weather event.

b) *Death Valley extreme heat (July 9-11, 2021; data snapshot from July 10, 2021)*: Following our analysis of the Pacific Northwest heat wave, we examined another heat wave centered on Death Valley. Using the seismic color palette for TMMX visualization as shown in Figure 23, we observed distinct patterns of extreme heat concentrated in

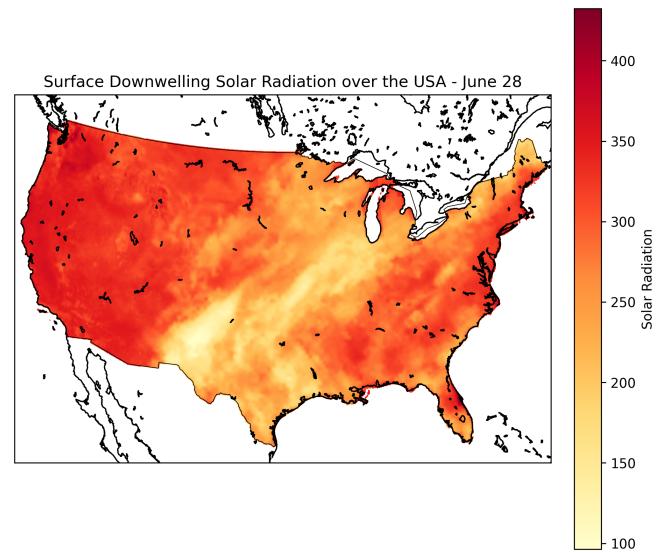


Fig. 22. June 28th - Surface downwelling shortwave radiation map using SRAD data

California's Death Valley. The YlOrRd visualization of solar radiation during this period as seen in Figure 24 showed intense radiation patterns that aligned with the temperature extremes, particularly in the Death Valley basin, extending into southern Nevada. This correlation helped establish the connection between solar radiation and temperature maxima in arid regions.

c) *Dixie Fire Development (July 13 - October 25, 2021; data snapshot from July 15, 2021)* [4]: The Dixie Fire, which became one of California's most destructive wildfires, was visualized by employing three distinct visualizations to understand the conditions that contributed to its rapid spread:

- Energy Release Component (ERC) visualization using the

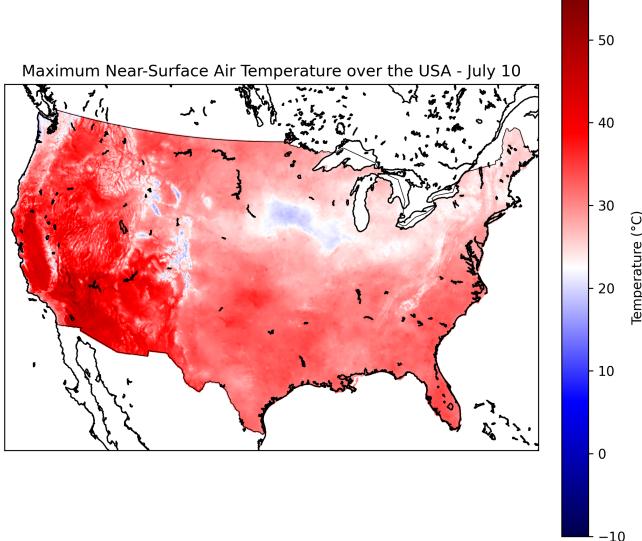


Fig. 23. July 10th - Maximum temperature map using TMMX data

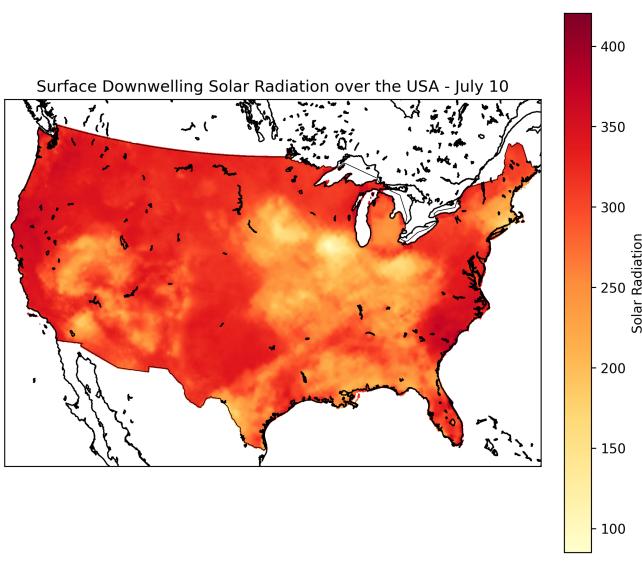


Fig. 24. July 10th - Surface downwelling shortwave radiation map using SRAD data

YlOrRd palette as shown in Figure 25 revealed areas of high potential fire intensity, with particularly severe conditions concentrated in the northern Sierra Nevada region and Plumas National Forest area.

- Burning Index (BI) mapping as shown in Figure 26, also using YlOrRd, showed a similar pattern of elevated fire danger extending through the Feather River Canyon and surrounding mountainous terrain, closely corresponding to the areas of high ERC values.
- 1000-hour Dead Fuel Moisture (FM1000) visualization using the BrBG diverging palette as seen in 27 highlighted critically low fuel moisture content in the affected region.

The combination of these three visualizations clearly demonstrates the dangerous confluence of conditions that led to the Dixie Fire's extreme behavior: high energy release potential, elevated fire danger, and extremely dry fuels all coincided in the same geographic area, creating ideal conditions for a major fire.

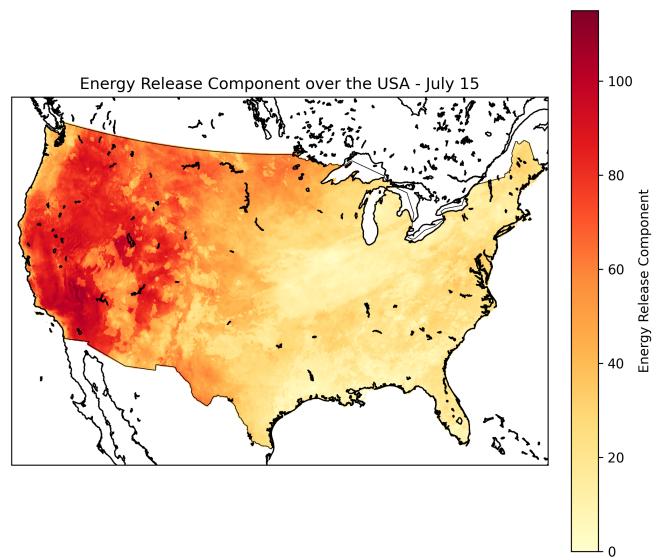


Fig. 25. July 15th - Energy Release Component map using ERC data

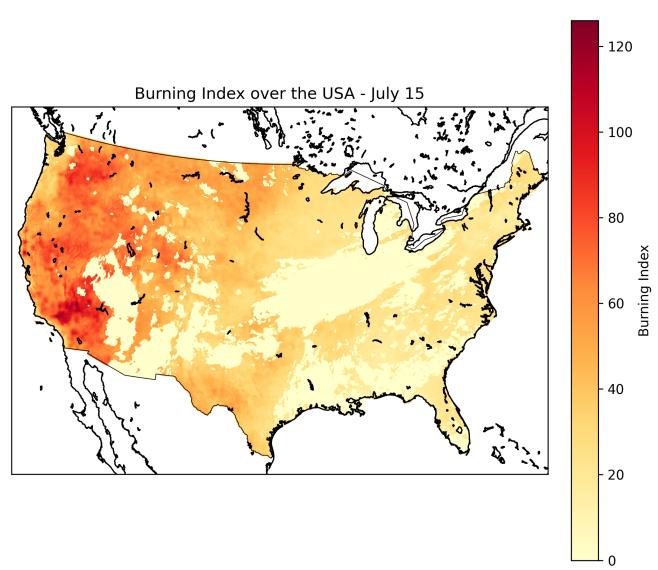


Fig. 26. July 15th - Burning Index map using BI data

d) Monument Fire Analysis (July 30 - October 7, 2021; data snapshot from August 2, 2021) [5]: Similar to the Dixie Fire analysis, we examined the Monument Fire using three visualizations. The patterns revealed by our color mapping techniques showed remarkable consistency across visualizations:

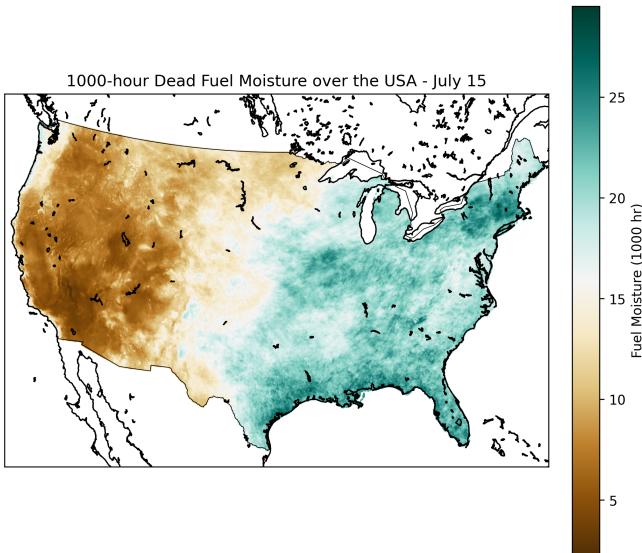


Fig. 27. July 15th - Dead Fuel Moisture (1000 hour) map using FM1000 data

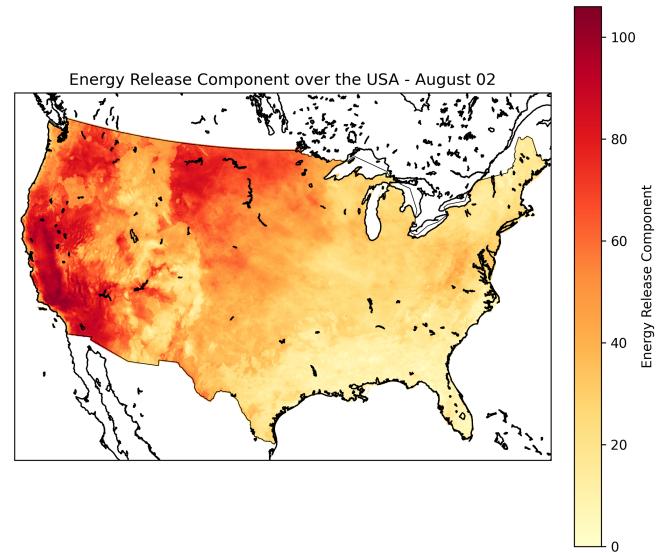


Fig. 28. August 2nd - Energy Release Component map using ERC data

- ERC values (YlOrRd palette) as seen in Figure 28 revealed intense fire potential concentrated in the Trinity Alps Wilderness area, with particularly severe conditions in the steep river canyons.
- High BI values (using YlOrRd palette) as seen in Figure 29 showed a nearly identical spatial pattern, with the most extreme fire danger indices aligned along the Trinity River corridor and surrounding mountainous terrain, validating the ERC findings.
- FM1000 visualization (using BrBGr palette) as seen in Figure 30 showing critically low fuel moisture content, particularly in the Trinity Alps region where the fire was most active

These three visualizations demonstrates why the Trinity Alps region was particularly susceptible to a significant fire event: the convergence of critically low fuel moisture, extreme energy release potential, and severe fire danger indices created optimal conditions for substantial wildfire activity.

B. Scalar Field Visualization using Contour Plots

The period between June 1st 2021 and August 31st was characterized by a number of weather events and catastrophic natural phenomena. Through contour plots we seek to identify these events by assessing conditions on days that preceded and succeeded them. Before doing so, we state the various experiments that were conducted to make the contours effective.

1) *Implementation and Experiments:* Contours are iso-lines that connect points that have the same value of a variable. However, these needn't always be concentric or form closed loops. Hence, making the visualizations appealing and interpretable becomes quite challenging. The color map used to fill between the contours, the number of contour levels shown in the map and the aspect ratios etc. turn out to be of utmost importance while conveying information. In this section we

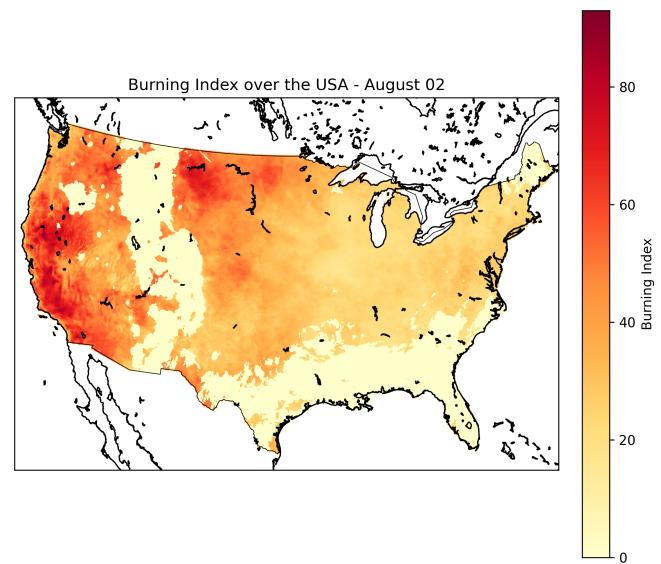


Fig. 29. August 2nd - Burning Index map using BI data

explain a few varieties of contour maps and provide reasons for choosing certain maps over others.

- 1) Plain Contours without text labels near each colored contour line are shown in Figure 31. The contours by default are generated using the marching squares algorithm. The text labels create unnecessary clutter and are hence avoided. These are quite informative for a single variable. However, these have some drawbacks. Increasing the number of levels can lead to excessive clutter. On the other hand, too few levels leads to a large amount of white space which leads to confusion about values at points within those regions.
- 2) Now we explore the use of colors as well as colored

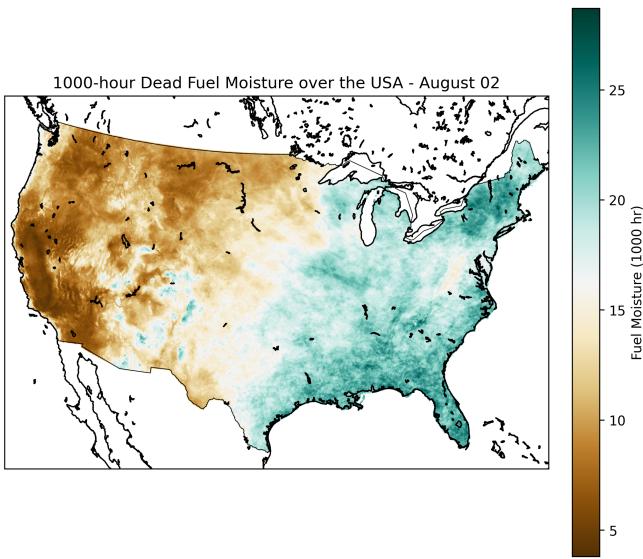


Fig. 30. August 2nd - Dead Fuel Moisture (1000 hour) map using FM1000 data

contour lines in Figure 32. The contour fill algorithm fills up grid cells that lie in between two contour lines with a single color. The contours are completely hidden and the map appears to be like any standard heatmap. To avoid this we suggest coloring the lines black and filling the regions between the contour lines with appropriate colors as done in Figure 34.

- 3) Contour Fills which use different colormaps were also explored. These use the contour fill algorithm to fill regions between contours with color. The best colormaps turned out to be the jet and jet_r colormaps despite the claim that "they introduce artifacts". The jet colormap is extremely intuitive for variables such as vapor pressure deficit, downward shortwave radiation and relative humidity. Dry/hot areas are shown in warmer colors like red while wet/cool areas are shown in cooler colors like blue. An example is shown in Figure 38. Since the focus is largely on whether variables cause dryness/heat or wetness/cold, we will use colormaps such as jet, coolwarm and spectral to encode these colors correctly.
- 4) We also explored visualizing two variables at once. This is done, for instance, by overlaying precipitation contours over a contour fill plot of humidity. Figure 33 shows overlaying downward surface radiation contours on contour fills of vapor pressure deficit. The problem with such plots is that they tend to be confusing as there is too much to process at once. Hence, we refrain from using this method to analyze two variables at once. Instead we can use juxtaposed views to compare two variables at once, whenever needed.
- 5) Last but not the least, using subtle parameters like power norms avoids concentration of data in a small portion of the colormap. Alternatively, logarithmic scales may also be used.

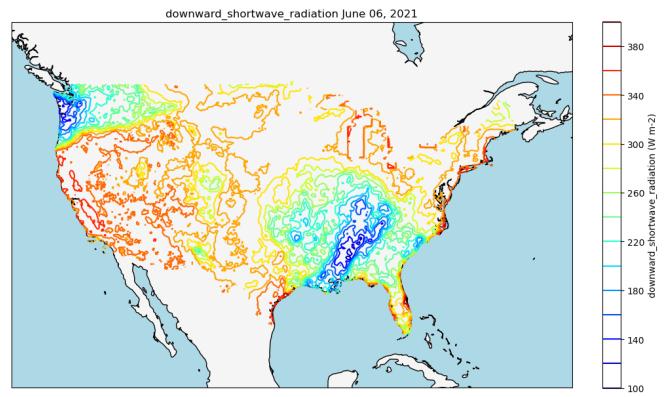


Fig. 31. A nice contour map showing contours for Downward Shortwave Radiation

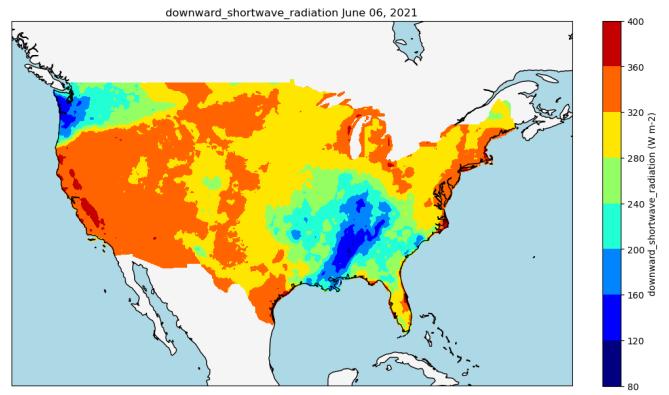


Fig. 32. Colored contour lines with filled regions in between. This appears like a colormap and ruins the point of using contours.

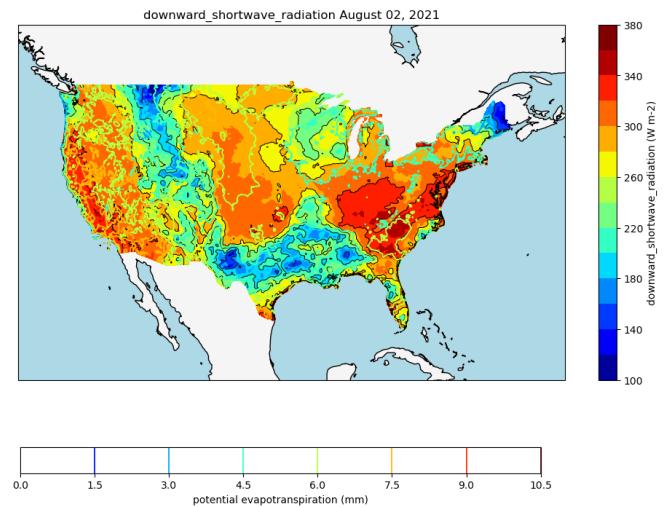


Fig. 33. Overlaying contours of one variable(downward surface radiation) over contour fills of another(vapour pressure deficit)

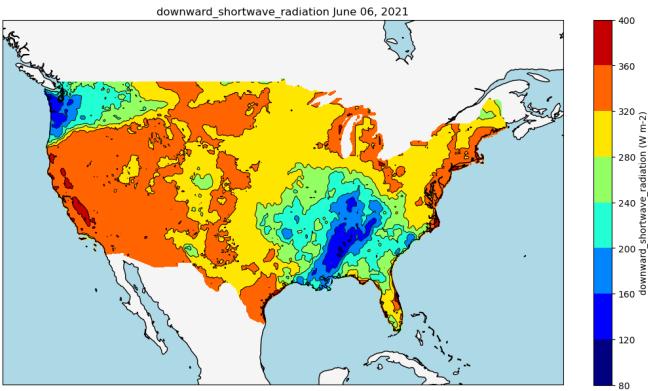


Fig. 34. Black contours with filled regions in between. This is the best so far.

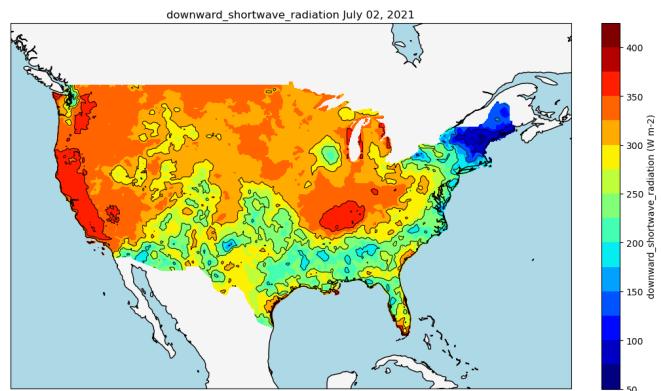


Fig. 37. July 2 2021. Wildfires in the west due to very high downward shortwave radiation and heat

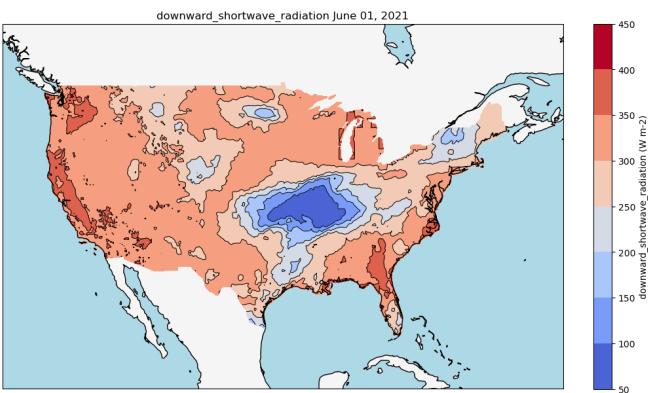


Fig. 35. July 01 2021. Downward Shortwave Radiation. See north western states like Oregon and Washington.

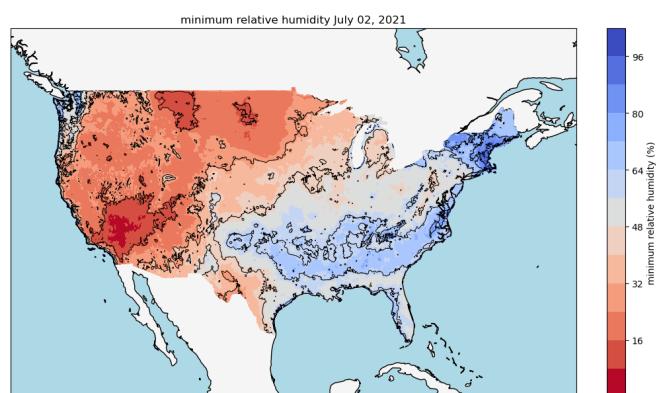


Fig. 38. July 2 2021. Wildfires in the west due to very low minimum relative humidity (shown in red)

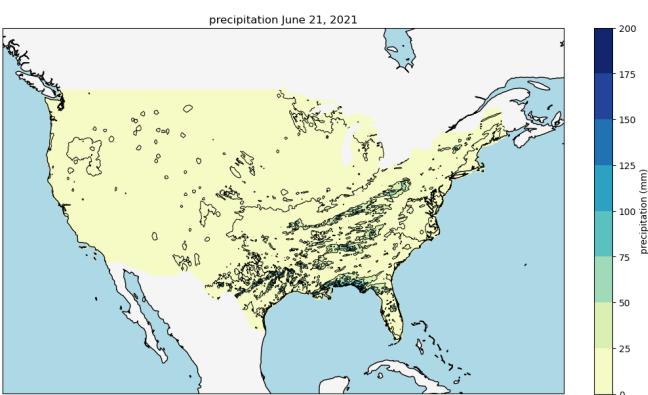


Fig. 36. June 21 2021. Flash flooding in parts of Texas, Louisiana and Mississippi. Most of the country was dry in this month (shown in yellow).

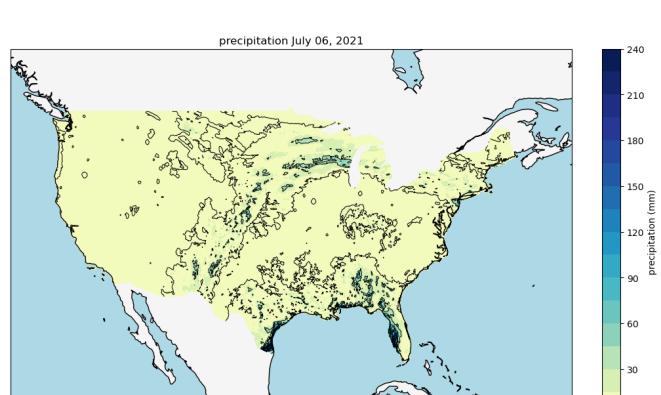


Fig. 39. July 6 2021. Precipitation in Florida due to hurricane Elsa

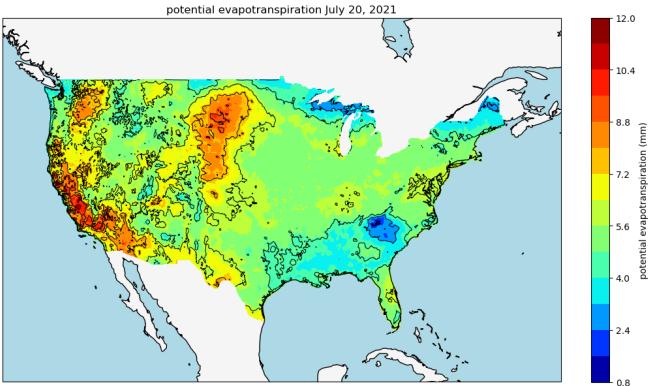


Fig. 40. July 20 2021. High Potential Evapotranspiration leads to drought, dryness and increased fire risk in the west

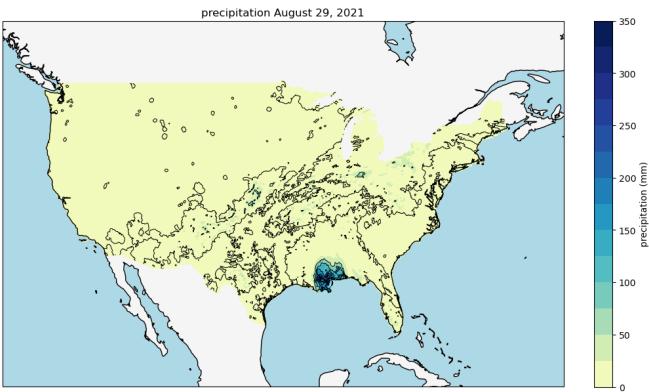


Fig. 41. August 29 2021. Hurricane Ida strikes Louisiana

2) Actual Visualizations and Inferences: In this section we will go through a list of natural phenomena that occurred in the period between June 1 2021 and August 31 2021 in the United States of America. Through contour plots we seek to understand the conditions that were prevalent at the time of these events. In the first week of June, extreme heat waves were observed in northwestern states of the country. Oregon and Washington received extremely high amounts of downward shortwave radiations which led to peak temperatures being recorded in Portland and Seattle respectively [6]. Figure 35 illustrates the previous statement.

The rest of June remained rather uneventful other than a few select thunderstorms and some extra heat in the southwest from June 15 to June 18. June 21 saw flash flooding [7] in parts of Texas, Louisiana, and Mississippi as shown in Figure 36. Here we use the yellow-to-blue colormap to avoid confusing dryness or lack of precipitation with heat. The rest of the U.S. received very low precipitation in Jun 2021 (Fig. 36). From July 1-5, wildfires in California, Oregon, and Washington started to intensify. The Dixie Fire in California became one of the largest active fires in the U.S. at the time, burning more than 200,000 acres. This can be seen in Figure 38. On July 6 2024, hurricane Elsa struck Florida. This can be seen in the small blue patches in the south east parts of the country (Fig.

39). The rest of July has other natural phenomena such as flash flooding in Kentucky from July 18-20, and some high temperatures in the North-East. More notably, the drought monitor reported that extreme drought (D3) conditions were widespread across the western U.S., particularly in California, Nevada, Oregon, and Utah [8]. This can be seen by visualizing the Potential Evapotranspiration in these areas. It is extremely high in dry, hot areas and causes plants to dry up. This in turn can lead to fires as dried plants tend to have more fuel mass. Figure 40 shows the Potential Evapotranspiration on July 20 2024. Now we turn to August. On August 4 and August 5, the Dixie fire in Northern California continued to rage, becoming the second-largest wildfire in California's history, with more than 650,000 acres burned by early August [9]. The contours for this are extremely similar to those in Figure 38 and are omitted to avoid unnecessary repetition. The most notable event in August was the occurrence of hurricane Ida [10]. On August 29, Hurricane Ida made landfall in Louisiana as a Category 4 storm with winds up to 150 mph (240 km/h). It caused catastrophic damage to New Orleans and surrounding areas, with widespread flooding and power outages, leaving over a million people without electricity. Figure 41 shows precipitation contours on August 29. Clearly, the blue parts peak at Louisiana, indicating heavy rainfall and drastic damage.

C. Vector Field Visualization using Quiver Plots and Streamlines

There was a notable storm system affecting the northwest part of the United States in the first week of June 2021 [11]. This period saw severe thunderstorms, heavy rainfall, hail, and strong winds across parts of the central and northern Plains. In addition to this, Hurricane Ida, the most damaging storm of the 2021 Atlantic hurricane season, made landfall in the southeastern part of the country on 29th August [10]. Therefore, we looked at 8 days split in two 4-day windows, 9-12 June and 28-31 August, to understand the impact of these storms via glyph-based vector field visualizations, namely quiver plots and streamlines.

The basemap library [12] was used for the visualizations. There are 3 types of visualizations - quiver plots with uniform sized arrows, where the color of the arrow represents the wind speed, another quiver plot with monochrome arrows and the length indicating speed, and a streamline plot which shows the wind flow, where speed is shown using color.

The rough sketch of the implementation is as follows:

- Select a suitable number of grid points for sampling the data, balancing resolution and computational efficiency. Index the speed and direction arrays appropriately and create a mesh grid.
- Extract the wind speed and direction for each point in the grid for a given day and encode it into a glyph (an arrow or a streamline). For the directions, the horizontal and vertical components are calculated separately.

- Create a basemap projection with continents, coastlines and countries as a background, and overlay the quiver plots or streamlines on top of it.
- For the streamlines, we also need to interpolate the wind speed and directions for points not on the grid, to obtain a continuous curve.

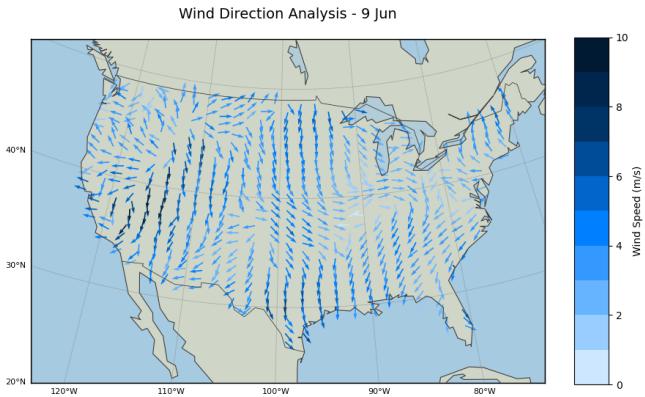


Fig. 42. Colored quiverplot for Jun 9

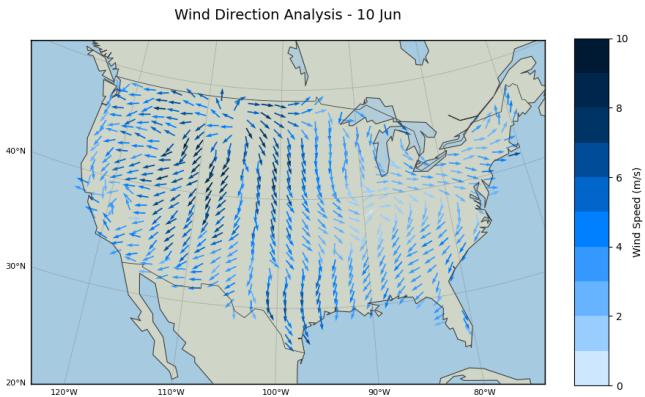


Fig. 43. Colored quiverplot for Jun 10. See diverging wind flow in the north

The following inferences can be drawn by observing the plots -

- In the northern part of the country, we can see winds flowing eastwards on 9th June (Figs.42,50,58), followed by high speed winds flowing southwards on June 10 (Figs.43,51,59). This was the peak of the storm. On the 11th (Figs.44,52,60) we see air from the east rushing to take the space left, and by 12th (Figs.45,53,61) we can see that the storm has died down and wind speeds have decreased. This can be seen in all 3 variants of the plots.
- In the month of August, a continuous southward wave of wind is observed on the 28th (Figs.46,54,62) near southeast USA, particularly the states of Louisiana, Alabama, Mississippi and Florida. We can see high speed eddies being formed on the 29th (Figs.47,55,63) and 30th

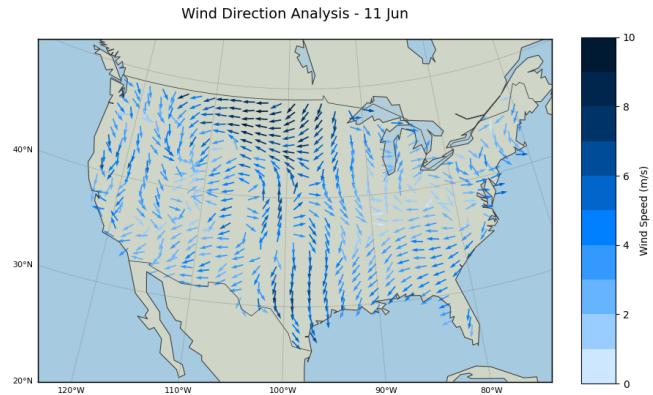


Fig. 44. Colored quiverplot for Jun 11. See high speed winds in the north

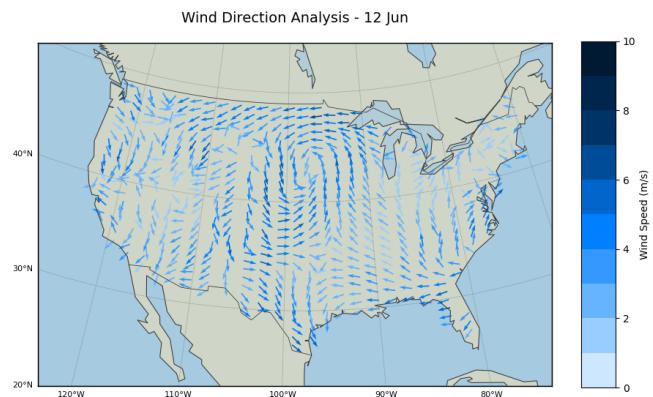


Fig. 45. Colored quiverplot for Jun 12

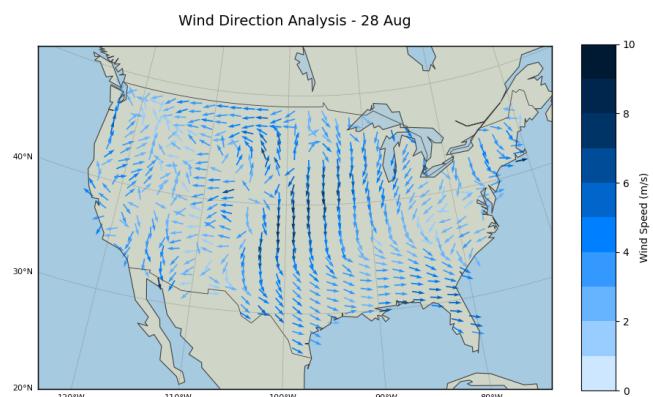


Fig. 46. Colored quiverplot for Aug 28

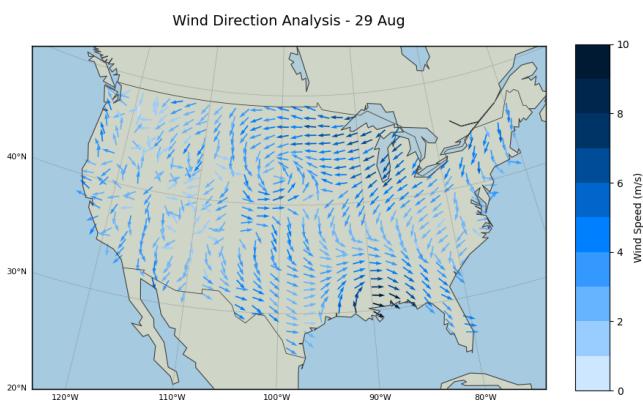


Fig. 47. Colored quiverplot for Aug 29

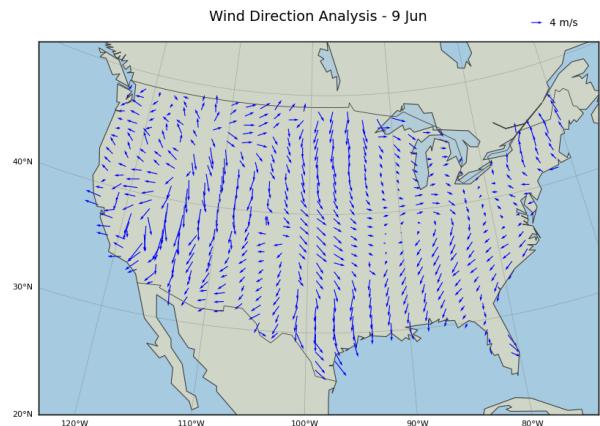


Fig. 50. Length-encoded quiverplot for Jun 9

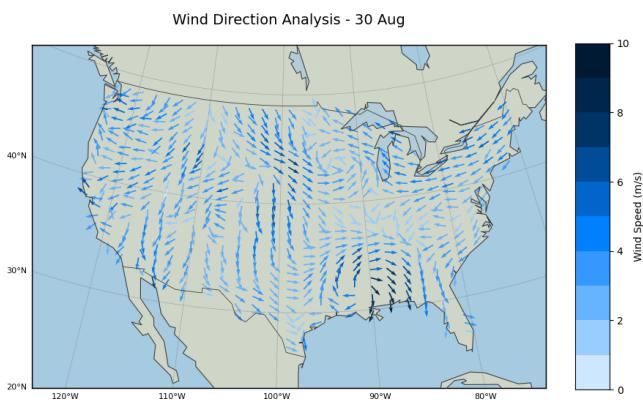


Fig. 48. Colored quiverplot for Aug 30. Fast winds in the southeast

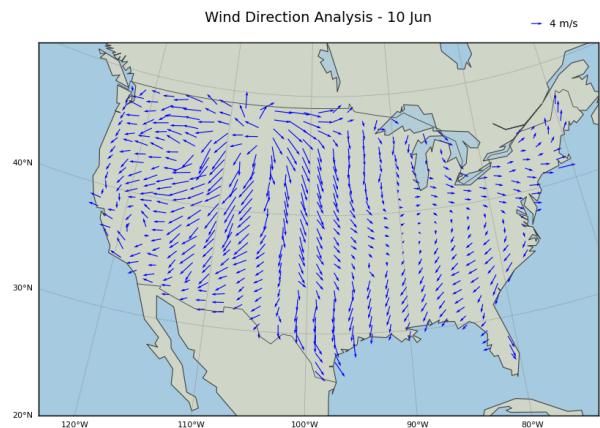


Fig. 51. Length-encoded quiverplot for Jun 10

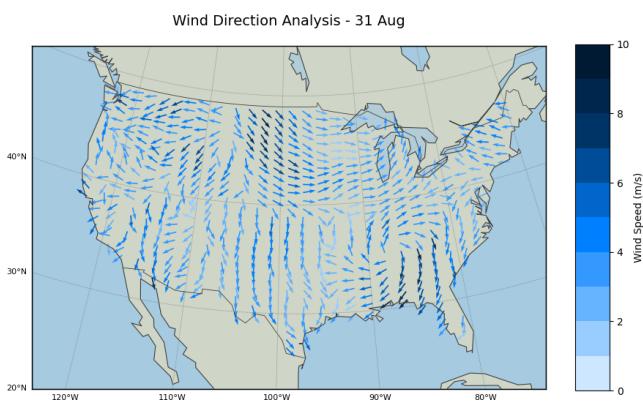


Fig. 49. Colored quiverplot for Aug 31

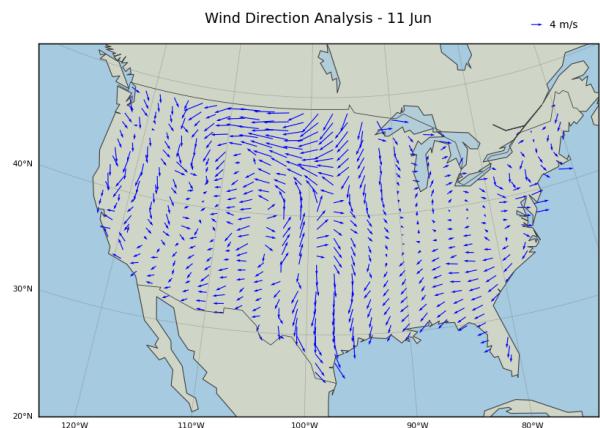


Fig. 52. Length-encoded quiverplot for Jun 11

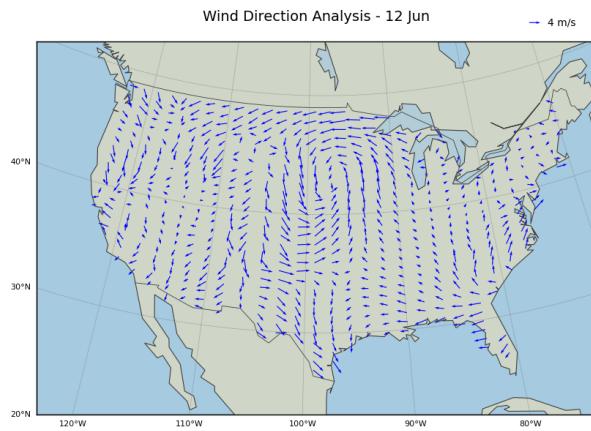


Fig. 53. Length-encoded quiverplot for Jun 12

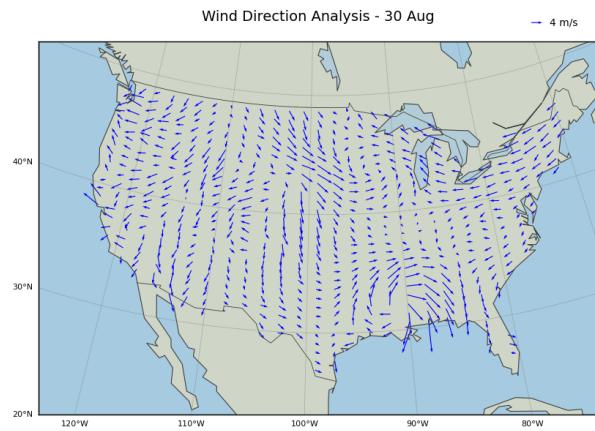


Fig. 56. Length-encoded quiverplot for Aug 30

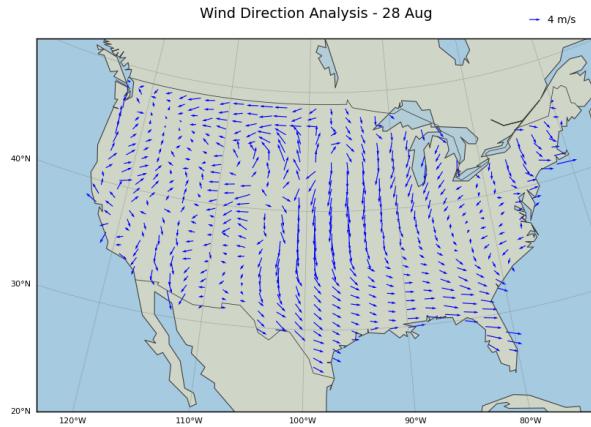


Fig. 54. Length-encoded quiverplot for Aug 28

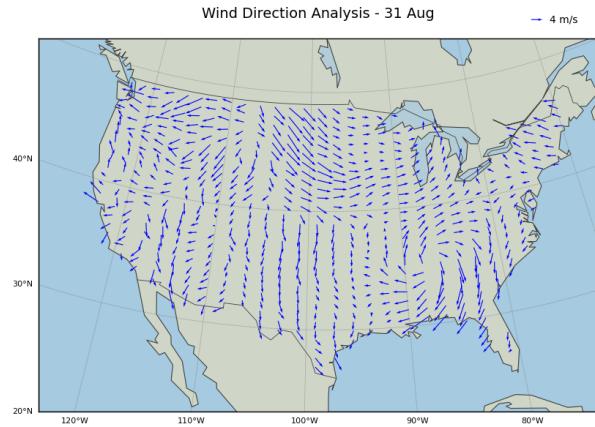


Fig. 57. Length-encoded quiverplot for Aug 31

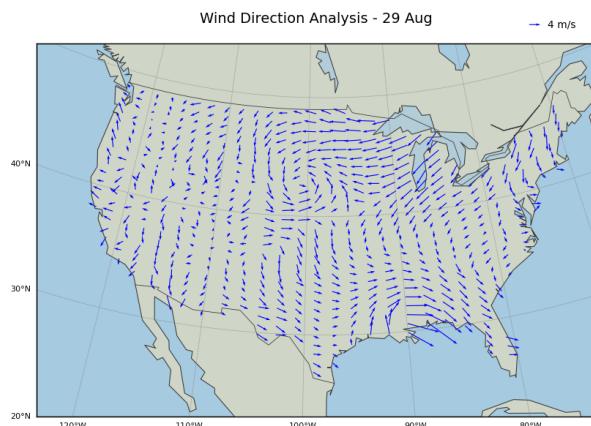


Fig. 55. Length-encoded quiverplot for Aug 29

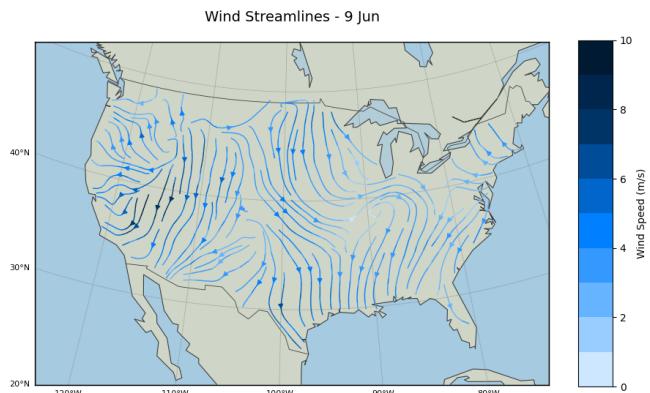


Fig. 58. Colored streamlines for Jun 9

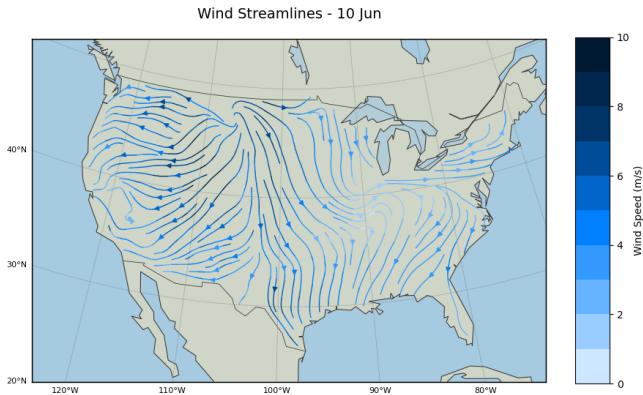


Fig. 59. Colored streamlines for Jun 10

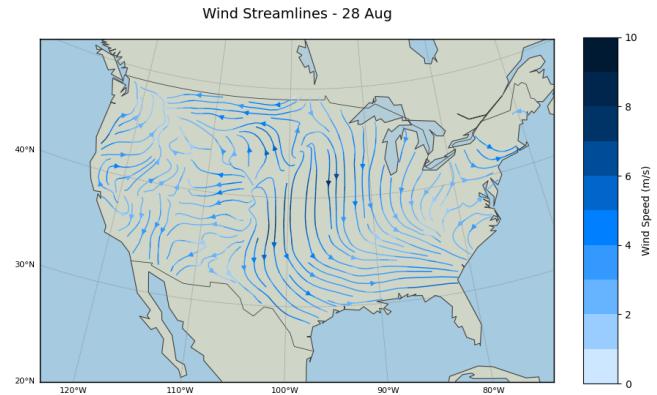


Fig. 62. Colored streamlines for Aug 28

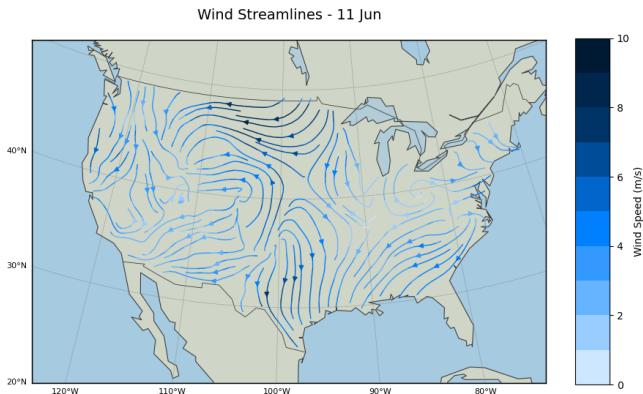


Fig. 60. Colored streamlines for Jun 11

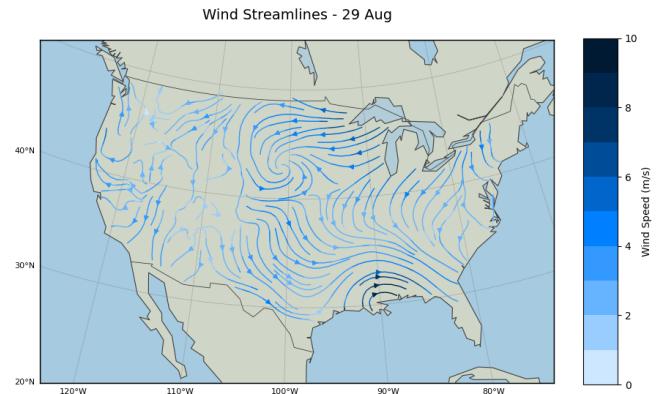


Fig. 63. Colored streamlines for Aug 29

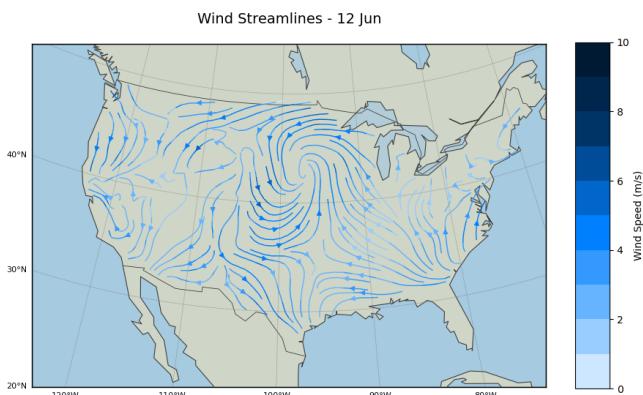


Fig. 61. Colored streamlines for Jun 12

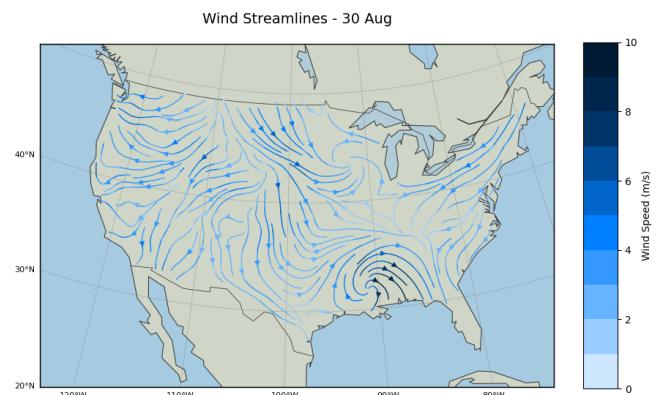


Fig. 64. Colored streamlines for Aug 30

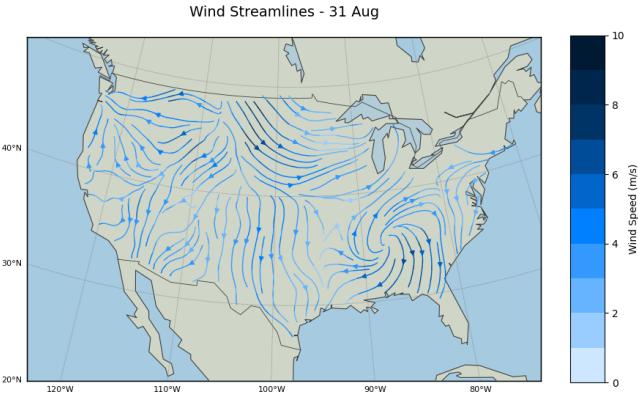


Fig. 65. Colored streamlines for Aug 31

(Figs.48,56,64) in the streamline plots, which is when Hurricane Ida made landfall. On the 31st, the wind speeds have decreased but are still significant (Figs.49,57,65).

- We can see winds flowing to the left in the western part of the country throughout the entire time period. This is in contrast to the normal west-to-east wind flow over the U.S. due to the westerlies [13]. These east-to-west winds were largely caused by a blocking high-pressure system and a strong heat dome over the region, which disrupted the typical west-to-east wind flow. The 2021 summer was one of the hottest summers in the southwestern part of the country i.e. the states of California, Arizona and Nevada. This led to hot air from that region rising up, and cold air from the neighbouring regions rushing to fill that space, which led to a continuous stream of moderate-speed winds over the entire time period.

In conclusion, glyph-based visualizations such as quiver plots and streamlines provide an effective way to understand vector fields like wind over a given region and a given period of time.

INFORMATION VISUALIZATIONS

In this part, we look at the Chinese Buddhism Dataset, taken from Github [14], which contains social data about historical Chinese Buddhist figures and their relationships, spanning over almost 2000 years. It has around 18000 nodes and 33000 edges. The nodes represent individuals, with their year of birth and death, gender and the ruling dynasty at that time. The edges represent documented relationships between individuals, such as discipleship, mentorship, collaboration, or other significant connections.

A considerable amount of preprocessing was required for this dataset. There were a large number of duplicate nodes and edges, which had to be removed. The graph also had a large number of very small components (<50 nodes), which we removed for clarity, and to focus on the main network. The birth years of several nodes were 0, which were imputed based on the birth years of their neighbors. Finally, there were over

120 distinct dynasties, most of which were small with only up to 50 members. So we filtered the nodes to include only the 15 largest dynasties. The graph after preprocessing had a single connected component with around 14000 nodes and 21000 edges. We then selected suitable parts of the graph as needed for our visualizations.



Fig. 66. Color mapping for different dynasties

A. Node-Link Diagrams

Visualizing graphs with over 10000 nodes and edges using node-link diagrams requires a lot of computational power, and leads to a lot of clutter. For that reason, we looked at a portion of the graph which only contains nodes representing individuals born in the years 0-800 CE, and the edges between them. This subset was chosen because these years were crucial in the introduction, integration and growth of Buddhism within Chinese society. We observe several dynasties, key historical figures in those times, and their connections with each other. For all graphs, the nodes are colored by the dynasty which they belonged to, and scaled in size according to their degree, i.e. nodes with more connections are larger. We experimented with the following different graph layouts -

Fruchterman-Reingold Layout - Nodes are initially randomly placed within the drawing area. Each node exerts a repulsive force on others to prevent overlap, given by $F_{ij} = \frac{k^2}{d_{ij}}$, where k is a constant and d_{ij} is the distance between nodes i and j . Connected nodes exert attractive forces to bring them closer, defined by $F_{ij} = \frac{d_{ij}^2}{k}$. Nodes move according to these force laws, and their movement is controlled by a temperature parameter that gradually decreases, helping the system stabilize to a final layout.

We get a circular layout with more or less evenly spaced nodes (Fig.67). Generally higher degree nodes are closer to the center in this layout, but the graph was sparse, so higher

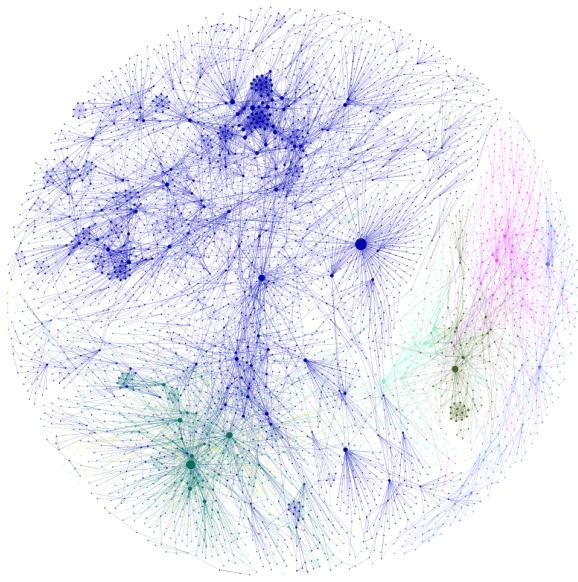


Fig. 67. The Fruchterman-Reingold Layout

degree nodes experienced uneven magnitudes of forces in different directions, which led them to move away from the center. Apart from easily identifying highly connected nodes, the Fruchterman-Reingold layout is ineffective in visualizing the network, as it does not make use of the inherently present temporal dimension. The algorithm applied on an ordered arrangement of the nodes also led to a similar final graph.

ing node positions, applying forces similar to Fruchterman-Reingold - repulsive forces between all nodes to avoid overlap and attractive forces between connected nodes to strengthen connections. Forces can be scaled by edge weights, allowing stronger connections to have more influence if edge weights are present. The Barnes-Hut approximation [15] is used to efficiently compute long-range repulsive forces in large graphs, reducing computational cost. A 'speed' parameter gradually reduces movement to allow convergence, while a 'gravity' parameter prevents nodes from drifting too far from the center [16]. For this algorithm, the initial layout was a single line of nodes on the X-axis ordered by their birth year, so the final output is more or less ordered by time as well.

The Force-Atlas layout offers a much more intuitive and easy-to-understand way of visualizing the network compared to the Fruchterman-Reingold layout, as we can see how it evolves over time (left to right) when the initial nodes are ordered (Fig.68). It starts with the Eastern Jin dynasty (light blue) in the first century, which is followed by the Liu-Song (pink), Southern Qi (olive) and Southern Liang (cyan) dynasties. In the 5th century, we have the Sui dynasty (green), which played a foundational role in shaping the trajectory of Buddhism in China, setting the stage for its further flourishing in the Tang dynasty (dark blue), often referred to as the 'Golden Age' of Chinese Buddhism [17]. During that period, Buddhism integrated deeply into Chinese society, culture and political life, which can easily be seen by the large and diverse blue network.

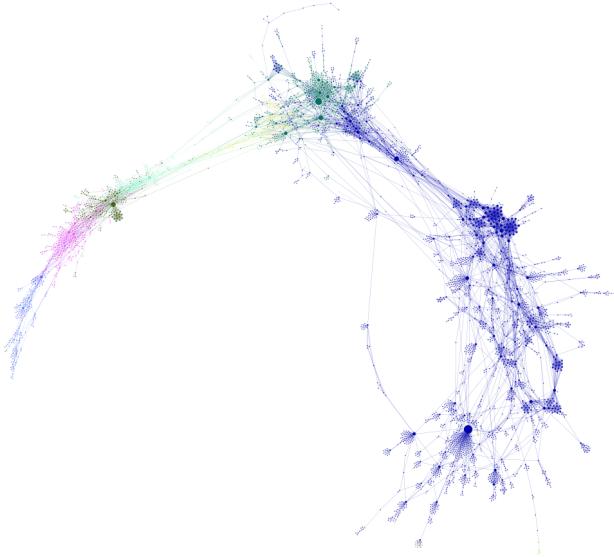


Fig. 68. The Force-Atlas 2 Layout

ForceAtlas 2 Layout - This layout starts with the exist-

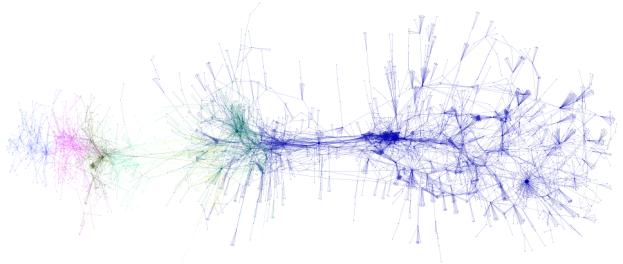


Fig. 69. The Hu Yifan Layout

Hu Yifan Layout - This layout is designed for efficient computation on large graphs [18]. It combines repulsive forces with an adaptive cooling schedule, allowing nodes to move quickly in the early stages and slowing as they approach

equilibrium. The repulsive forces are calculated using the Barnes-Hut approximation to handle long-range interactions efficiently. Additionally, attractive forces between connected nodes bring them closer, emphasizing local structure. Unlike Fruchterman-Reingold, Hu Yifan's layout scales the strength of repulsive and attractive forces adaptively based on node density, which helps prevent overlapping and improves layout clarity in dense regions of the graph. Here as well, we kept the initial layout as a linear arrangement of nodes in order of their year of birth (Fig.69).

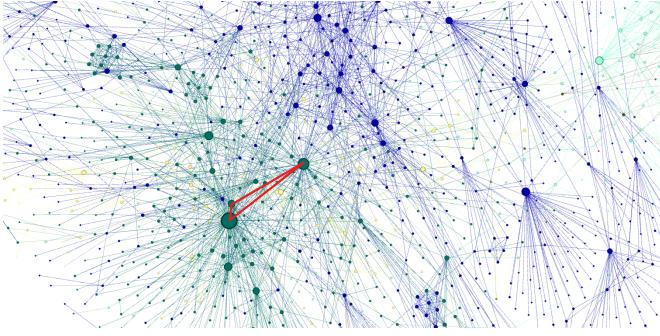


Fig. 70. Sui dynasty triangle, Fruchterman-Reingold Layout

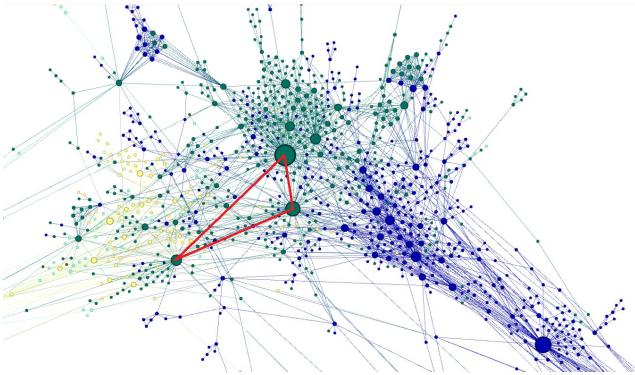


Fig. 71. Sui dynasty triangle, Force-Atlas Layout

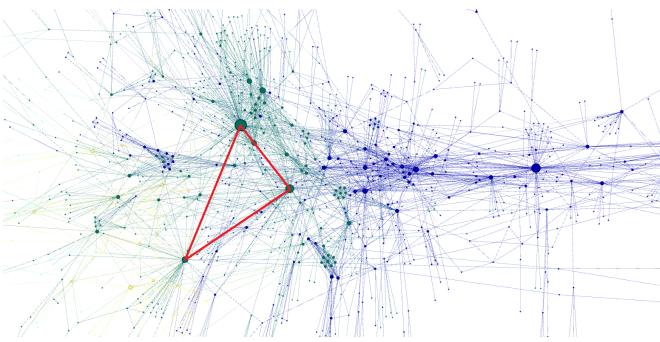


Fig. 72. Sui dynasty triangle, Hu Yifan Layout

The Yifan Hu layout puts nodes further apart than the ForceAtlas Layout. This makes it easier to view the network topology and identify individual nodes. In all three layouts,

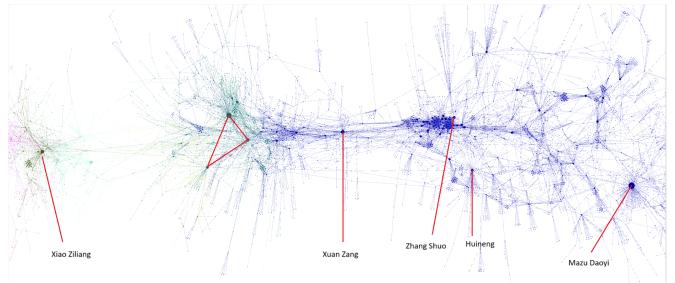


Fig. 73. Notable individuals marked in the Hu Yifan Layout

you can see a characteristic triangle in the Sui dynasty (Fig.70,71,72), the vertices of which are three notable Buddhist figures -

- Yang Jian - The founder of the Sui dynasty [19]. He unified China after a series of conflicts, and supported Buddhism in a time where Confucianism was prevalent. The Sui dynasty is known for its significant infrastructure projects, which also included building Buddhist Monasteries.
- Tan Qian - A Buddhist monk who is known for his contributions to the translation of Buddhist scriptures into Chinese during the Sui Dynasty [20]. His translations were important in spreading Buddhist teachings and philosophy in China, making them more accessible to Chinese audiences.
- Huiyuan - Founder of the Pure Land School of Buddhism in China [21]. He is credited with laying the foundation for what would later become one of the most popular forms of Chinese Buddhism, especially among laypeople.

You can also see Xuan Zang, the famous Buddhist scholar who travelled to India during the period of emperor Harsha [22], and Mazu Daoyi, the father of Zen Buddhism in China and Japan [23]. In addition to these, there are many other individuals who have had a significant contribution to making Chinese Buddhism what it is today (Fig.73). The Hu Yifan and Force-Atlas layouts were found to be highly effective in visualizing and communicating the importance and impact of these individuals through node-link diagrams.

B. Parallel Coordinate Plots

The primary objective of a parallel coordinates plot(PCP) is to visualize multiple variables or 'columns' at the same time in a single visualization. Any basic parallel coordinates plot consists of a number of vertical axes each representing a variable or attribute of a data point. Each data point is then drawn in the form of a line that goes through its attribute values corresponding to each column.

In this dataset, we present two different variants of parallel coordinate plots.

The first treats each individual person (a node in the dataset) as a poly-line or a single fundamental unit of the PCP. This PCP allows us to analyze the dataset at an individual level and answer questions of the type "Which individual has the

most number of interactions?".

The next one treats each nationality as a poly-line and the vertical axes represent aggregate attributes of the dynasties. This allows us to answer questions at a higher level such as "Which dynasty has the most number of interactions?"

Note that we do not answer questions on the count of individuals each dynasty as those are better seen in treemaps. Although we could simply add another axis corresponding to the count of individuals in a dynasty, we refrain from doing so as an increase in the number of axes in the PCP affects the ability to perceive and fully comprehend the poly-lines. More details on such aspects as well as the use of interactions are provided in the sections that follow.

1) Variant 1: Individual level focus: As stated earlier, here each poly-line represents a single individual i.e. his/her birth year, nationality and number of interactions. We have divided the dataset into three parts: 0-800, 801-1400 and 1400-end to avoid having to many categories in the nationality column and making the PCPs difficult to comprehend. This division also represents the different "eras" of Buddhism in some sense. Figure 74, 76 and 78 show such PCPs. Lines are colored by nationality and the points are ordered in a way that minimizes crossings.

Each PCP allows reordering of axes in order to view the relation between two variables closely. This can be done by clicking on the top of the axis (near the label) and dragging the axis left or right of its neighbors to determine the new ordering of axes. The other interaction that PCPs provide is brushing. Brushing refers to selecting a portion of the dataset. An example is shown in Figure 75. We brush the last axis to see the individuals with the most number of interactions. We see that two individuals from the Tang dynasty have the most number of interactions.

Similar figures are shown for the other parts i.e. birth year between 801 to 1400 and birth year exceeding 1400 (Fig. 76,78).

We explore another brushing operation here. In Figure 77, we have brushed the birth years between 1250 and 1350 and then selected every dynasty except the Ming nationality. This tells us which nationalities were active in the period between 1250 and 1350 other than the Ming dynasty. In similar ways PCPs can be used to answer a wide variety of questions.

Lastly, we present the reordering interaction in this variant which allows us to drag an axis over another to change the order of axes. Such a reordering is shown in Figure 79. This kind of reordering may be useful to clearly view the relationships between two variables(axes) by placing them close together. This variant provides only individual level data and can answer questions that start with the phrase "How many individuals ...?" or "Which individuals ...?".

2) Variant 2: Nationality level focus:: As stated earlier, here each poly-line represents a single nationality i.e. its average birth year and its number of interactions. We calculate the self interaction, cross interaction and total interaction ratios as follows:

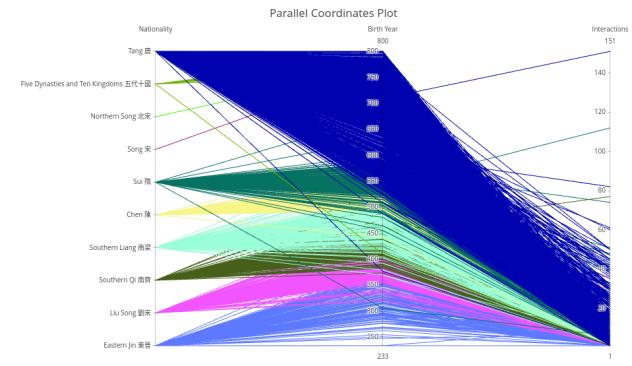


Fig. 74. PCP variant 1 for individuals with birth year between 0 and 800.

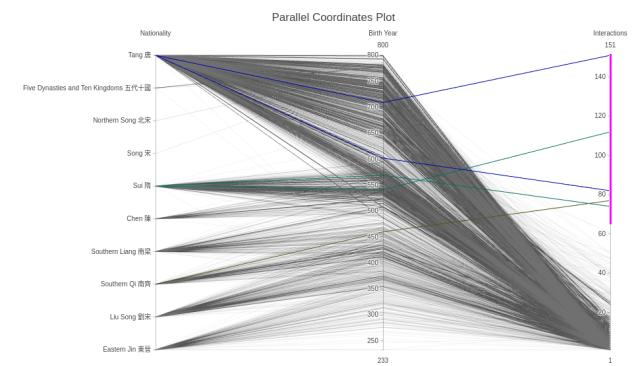


Fig. 75. PCP variant 1 for individuals with birth year between 0 and 800. Interactions axis is brushed to see only individuals with the highest number of interactions

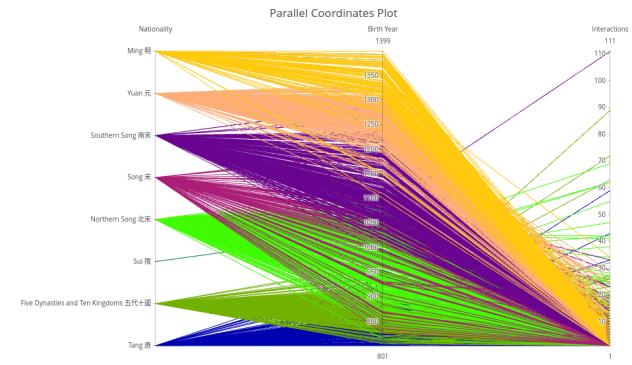


Fig. 76. PCP variant 1 for individuals with birth year between 801 and 1400.

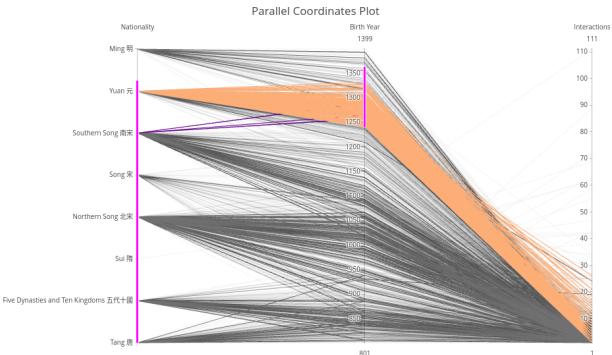


Fig. 77. PCP variant 1 for individuals with birth year between 801 and 1400. Brushing two axes to see the number of individuals that are not from the Ming Dynasty and are born in the period between 1250 and 1350

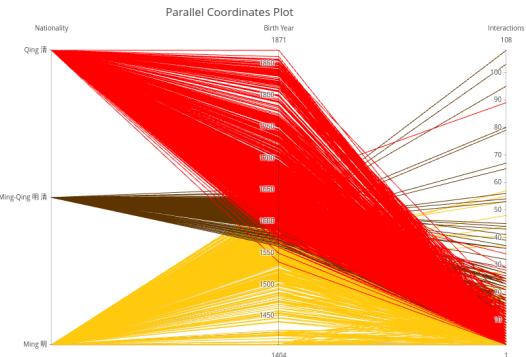


Fig. 78. PCP variant 1 for individuals with birth year more than 1400 years (till about 1896 years).

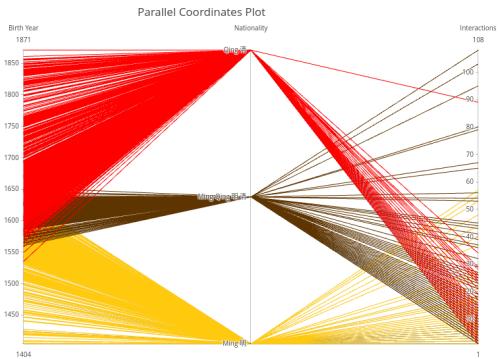


Fig. 79. PCP variant 1 reordering of axes to see clearer pairwise relationships).

self interaction ratio = $2 * \frac{\#\text{edges_to_nodes_of_same_nationality}}{\text{number of nodes}}$
 cross interactions ratio = $\frac{\#\text{edges_to_nodes_of_other_nationality}}{\text{number of nodes}}$
 total interactions ratio = self interaction ratio + cross interaction ratio but this does not include the factor of 2 that is multiplied above.

So, total interactions ratio isn't really a "total" value. Rather it is a metric that emphasizes the cross-interactions over self-interactions.

Note that dividing by the number of nodes in the denominator is to provide some sort of normalization. Interaction ratios should be a measure of how interactive each nationality is which is more dependent on the "nature" of the people rather than the "number" of people. Figures 80, 81 and 82 show these PCPs. Additionally, figure 83 provides a combination of the previous three figures.

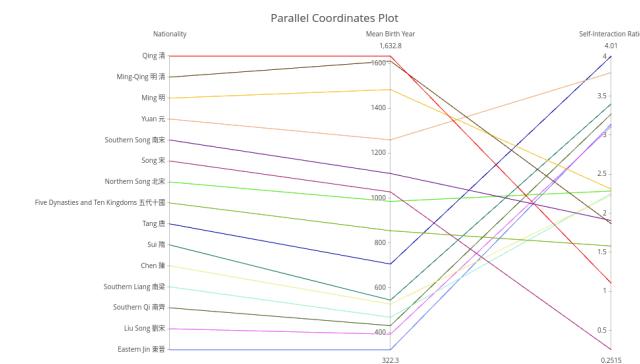


Fig. 80. PCP variant 2 self interaction ratio.

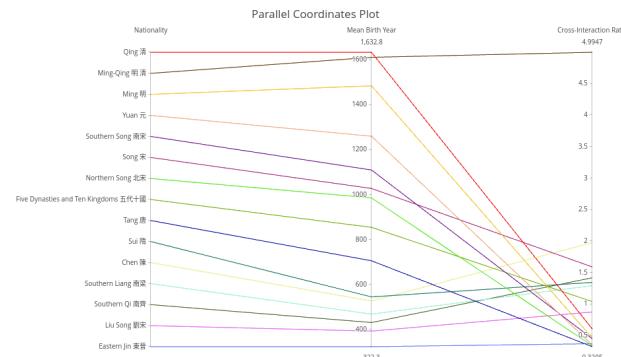


Fig. 81. PCP variant 2 cross interaction ratio.

Here there is little need for brushing as the dynasties/nationalities are few and the plot is not cluttered. The mean birth year is placed next to the nationality column to roughly show the period of time for which the dynasties made significant contributions to Buddhism. Other measures such as the median could also be used.

From Figure 80 we see that the Tang, Yuan, Sui and Chen dynasties have the most interactions among themselves. On the

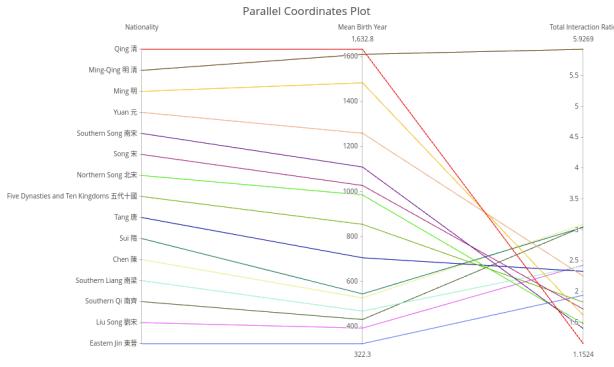


Fig. 82. PCP variant 2 total interaction ratio.

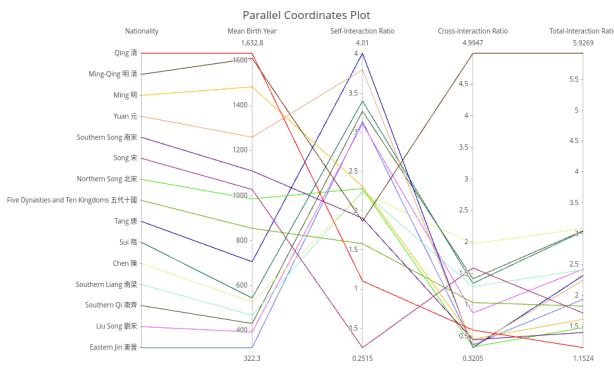


Fig. 83. PCP variant 2 all interaction ratios in a single plot.

other hand from Figure 81 we see that Ming-Qing nationality has a very significant cross interaction ratio when compared with any other nationality. In terms of total interactions(which is biased towards the cross interaction metric), the Ming-Qing dynasty has a huge head over any other dynasty. Next we see the Chen dynasty which performs pretty well in both self and cross interaction ratio measures. The Sui dynasty comes in at third place due its large number of cross interactions.

Thus PCPs prove to be extremely useful while analyzing various interaction measures at the dynasty level.

C. Treemaps

The main goal of creating treemap visualizations was to reshape raw data into a meaningful tree structure to capture key trends across the dataset. By doing so, we could explore different ways of organizing and visualizing data hierarchically, which helped reveal connections within historical and social contexts. Three treemap visualizations were created, each highlighting distinct hierarchical aspects to offer unique perspectives on these relationships. In each of the three visualizations, experiments have been performed with four different-layout algorithms : Squarify, Binary, Slice and Dice. The treemap visualizations include hover functionality to allow users to view node names that may not be clearly visible on the rectangles.

1) Treemap 1: Dynasty and Gender: In this treemap, the data is initially displayed as a distribution of nodes by dynasty, allowing viewers to see the representation of each dynasty within the dataset. The top-level in Fig. 84 shows that Qing (29.1%) and Tang (21.2%) are the two major dynasties to which more than half of the nodes belong to. Upon clicking a specific dynasty, the treemap further subdivides to reveal the gender distribution within that group as shown in Fig. 85. This drill-down feature provides a more interactive experience, enabling an in-depth look at gender representation across different dynasties. From this visualization, it is evident that female nodes are very few in most dynasties, reflecting a male-dominated, patriarchal structure within the dataset and suggesting historical biases in gender representation.

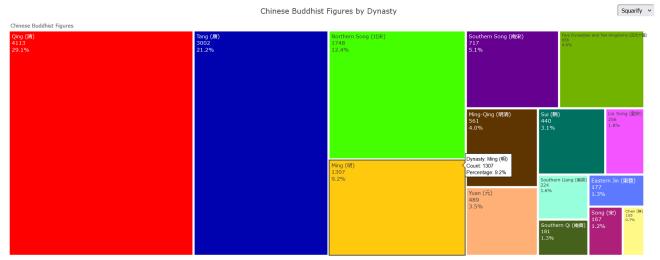


Fig. 84. Treemap 1 - Top-level view using Squarify algorithm for layout



Fig. 85. Treemap 1 - User Interaction: after clicking on 'Liu Song' dynasty

2) Treemap 2: Time Period and Nationality: In this treemap, the data is first displayed as a distribution of nodes based on birth years, grouped into three historical time periods: 0–800 CE, 801–1400 CE, and 1401–2000 CE. This layout as seen in Fig.86 allows viewers to observe how Chinese Buddhist figures are distributed across these eras. The top-level view highlights the concentration of records within each period, revealing a progressive increase in the number of figures documented over time, with the fewest figures from 0–800 CE, more from 801–1400 CE, and the highest count from 1401–2000 CE. This increase in the number of Chinese Buddhist figures over time reflects the growing spread and influence of Buddhism in China with time. Upon selecting a specific time period, the treemap further subdivides to display the distribution of dynasties within that era as shown in Fig. 87. This allows viewers to explore the representation of various dynasties across the three historical periods. Within each time

period, one dynasty is especially prominent: Tang for 0–800 CE, Song (Northern Song, Southern Song and Song) for 801–1400 CE, and Qing for 1401–2000 CE. The number and distribution of dynasties within each time period indicates the longevity and historical presence of specific dynasties.

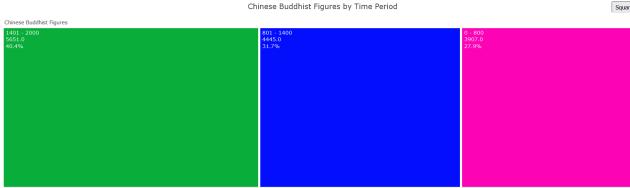


Fig. 86. Treemap 2 - Top-level view using Squarify algorithm for layout

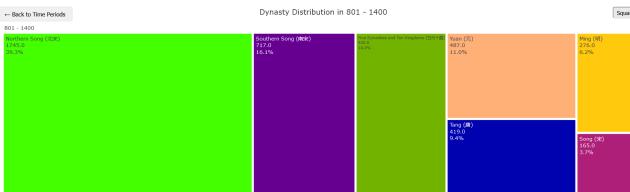


Fig. 87. Treemap 2 - User Interaction : after clicking on '801-1400' block

3) Treemap 3: Interaction Counts and Interactions with Specific Dynasties: Interactions are calculated by the number of times a node appears as a source or target node for an edge. In the top-level view, the nodes are grouped by dynasties and the value associated with a dynasty is the number of its interactions as shown in Fig. 88. One interesting observation here is that even though Qing is the most prominent dynasty by number of nodes as seen in Fig. 84, Tang is the dynasty with the most interactions making it the most influential dynasty. On clicking a dynasty and clicking the box again, we can see the distribution of interactions of that dynasty including self-interactions. Most of the dynasties have the number of self-interactions as the prominent group in that level but an exception to this is the Qing dynasty which has maximum interactions with Buddhist figures from the Ming-Qing dynasty as seen in 89.

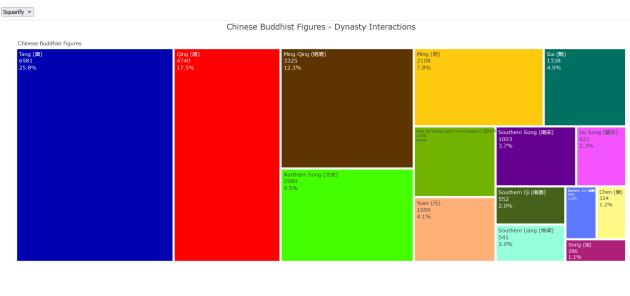


Fig. 88. Treemap 3 - Top-level view using Squarify algorithm for layout

For the top-level view of first treemap, the Squarify algorithm is the best choice as it produces a proportional layout

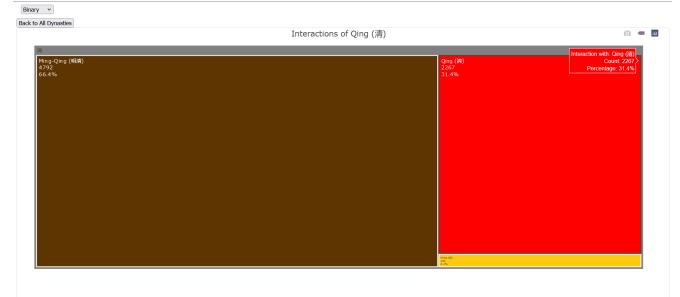


Fig. 89. Treemap 3 - User Interaction : after clicking on 'Qing' dynasty twice

that clearly presents the dynasty distributions. While Squarify may sacrifice some compactness, the improved readability and balanced representation of large and small categories makes it the preferred option. For the second-level of the treemap showing gender distribution in a dynasty, any layout is fine as we have only two groups - Male and Female. In the second treemap top-level view showing distribution of nodes across time periods, the layout algorithm has less of an impact, as Squarify, Binary, and Dice all generate similar, balanced views on the three time periods. However, the Slice algorithm diverges by stacking time periods linearly, which causes visual fragmentation and reduces clarity. For the top-level view of the third treemap, both Squarify and Binary are strong options. Squarify provides a visually appealing and intuitive layout. While Binary maintains a clear, hierarchical structure that makes it easy to compare interaction counts between dynasties and identify dominant connections, it results in a less intuitive ordering of rectangle sizes compared to Squarify. Given this, Squarify may be the best overall choice for the third treemap as well, balancing interpretability and aesthetics.

Figures 90, 91, 92, 93, 94, 95 and 96 show different layouts in the top-level of each of the three treemaps other than Squarify which was shown before. Overall, we can see that Squarify is the best algorithm for all three treemaps, followed closely by Binary in second. The Dice and Slice algorithms, which stack groups horizontally and vertically respectively, do not use a compact layout with intuitive ordering of the rectangle sizes.

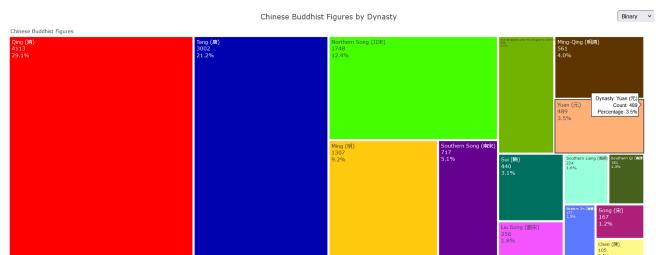


Fig. 90. Treemap 1 - Top-level view using Binary algorithm for layout

AUTHORS' CONTRIBUTIONS

- IMT2022017 Prateek Rath - Contour Maps and Parallel Coordinates Plots

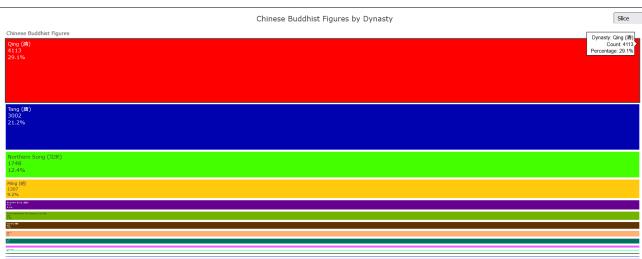


Fig. 91. Treemap 1 - Top-level view using Slice algorithm for layout

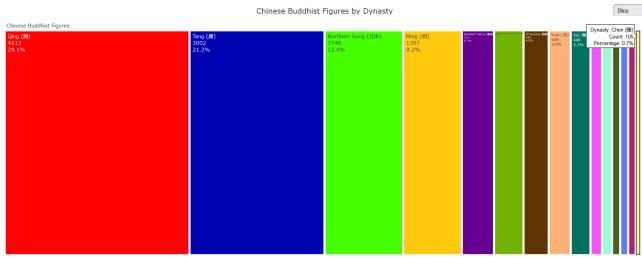


Fig. 92. Treemap 1 - Top-level view using Dice algorithm for layout

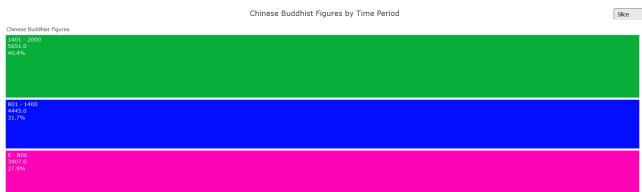


Fig. 93. Treemap 2 - Top-level view using Slice algorithm for layout

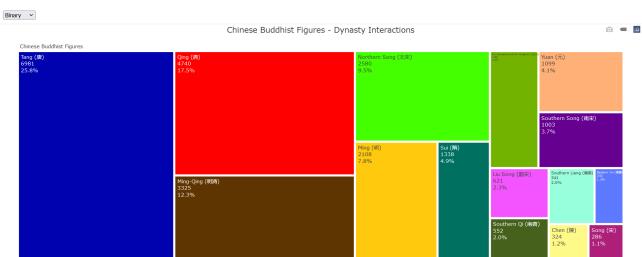


Fig. 94. Treemap 3 - Top-level view using Binary algorithm for layout

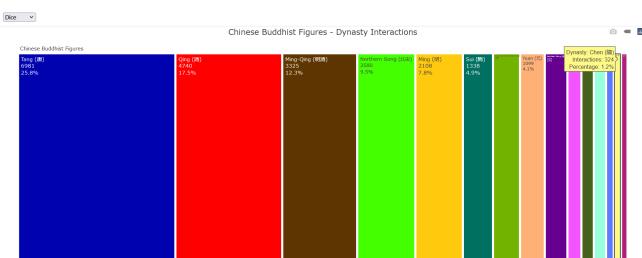


Fig. 95. Treemap 3 - Top-level view using Dice algorithm for layout

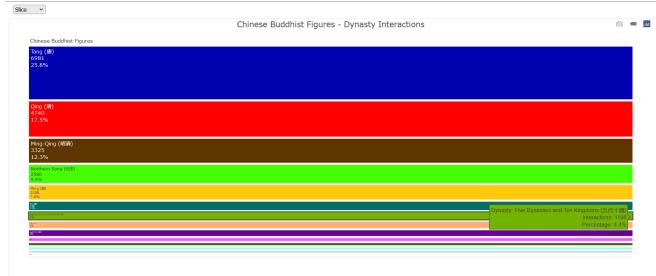


Fig. 96. Treemap 3 - Top-level view using Slice algorithm for layout

- IMT2022076 Mohit Naik - Quiver Plots and Node-Link Diagrams

- IMT2022103 Anurag Ramaswamy - Colormaps and Treemaps

Each team member worked independently on their assigned task, including all aspects of data analysis, visualization creation, interpretation, and description. The preprocessing for the information vizualizations was done jointly by all three members.

REFERENCES

- [1] C. Lab, "Gridmet: Gridded meteorological data." [Online]. Available: <https://www.climatologylab.org/gridmet.html>
- [2] NOAA National Centers for Environmental Information, "August 2021 national climate report." [Online]. Available: <https://www.ncei.noaa.gov/access/monitoring/monthly-report/national/202108>
- [3] W. contributors, "2021 western north america heat wave." [Online]. Available: https://en.wikipedia.org/wiki/2021_Western_North_America_heat_wave
- [4] U.S. National Park Service, "Dixie fire." [Online]. Available: <https://www.nps.gov/lavo/learn/nature/dixie-fire.htm>
- [5] California Department of Forestry and Fire Protection, "Monument fire." [Online]. Available: <https://www.fire.ca.gov/incidents/2021/7/30/monument-fire>
- [6] TimeAndDate, "Weather in june 2021 portland." [Online]. Available: <https://www.timeanddate.com/weather/usa/portland/or/historic?month=6&year=2021>
- [7] Floodlist, "flood list louisiana." [Online]. Available: <https://floodlist.com/tag/louisiana>
- [8] NCEI, "drought." [Online]. Available: <https://www.ncei.noaa.gov/access/monitoring/monthly-report/drought/202107>
- [9] NPR, "more fire in california." [Online]. Available: <https://www.npr.org/2024/08/04/g-s1-15279/firefighters-continue-battling-massive-wildfire-in-california>
- [10] National Hurricane Center, "Tropical cyclone report: Hurricane ida (al092021)." [Online]. Available: https://www.nhc.noaa.gov/data/tcr/AL092021_Ida.pdf
- [11] The Washington Post, "Weather disasters in 2021: A year of extremes." [Online]. Available: <https://www.washingtonpost.com/nation/interactive/2021/weather-disasters-2021/>
- [12] M. Contributors, "Basemap toolkit for matplotlib." [Online]. Available: <https://matplotlib.org/basemap/stable/>
- [13] W. contributors, "Westerlies." [Online]. Available: <https://en.wikipedia.org/wiki/Westerlies>
- [14] M. Bingenheimer, "Historical social network of chinese buddhism." [Online]. Available: https://github.com/mbingenheimer/ChineseBuddhism_SNA
- [15] A. Contributors, "Barnes-hut approximation in arbor.js documentation." [Online]. Available: <http://arborjs.org/docs/barnes-hut>
- [16] P. ONE, "Analysis of the structure and dynamics of the internet backbone." [Online]. Available: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0098679>
- [17] W. contributors, "Sui dynasty." [Online]. Available: https://en.wikipedia.org/wiki/Sui_dynasty

- [18] Y. Hu, "Graph drawing: Small graphs." [Online]. Available: http://yifanhu.net/PUB/graph_draw_small.pdf
- [19] W. contributors, "Emperor wen of sui." [Online]. Available: https://en.wikipedia.org/wiki/Emperor_Wen_of_Sui
- [20] U. U.-C. Institute, "Jinhua monks and monarchs: Kinship and kingship, tanqian, sui buddhism, and politics." [Online]. Available: <https://china.usc.edu>
- [21] W. contributors, "Lushan huiyuan." [Online]. Available: https://en.wikipedia.org/wiki/Lushan_Huiyuan
- [22] E. B. Contributors, "Xuanzang," 2024. [Online]. Available: <https://www.britannica.com/biography/Xuanzang>
- [23] B. Now, "A taste of zen: Mazu daoyi." [Online]. Available: <https://buddhismnow.com/2015/09/19/a-taste-of-zen-mazu-daoyi/>