Data Visualization : Assignment 2

Ayush Arya Kashyap
IMT2022129
IIIT - Bangalore
Bangalore, India
ayusharya.kashyap@iiitb.ac.in

Uttam Hamsaraj
IMT2022524
IIIT - Bangalore
Bangalore, India
uttam.hamsaraj@iiitb.ac.in

Pranav Laddhad
IMT2022074
IIIT - Bangalore
Bangalore, India
pranav.laddhad@iiitb.ac.in

I. Scientific Visualization

A. Quiver Plotting

A quiver map is a type of plot that uses arrows to represent vector fields, displaying both the direction and magnitude of data at specific points. In meteorology and oceanography, for example, quiver maps show wind or ocean currents, where the arrow direction indicates the flow (wind direction) and the arrow color or length represents the speed or strength. This visualization helps in quickly understanding the patterns and intensity within a vector field over a geographical or grid-based area.

1) *DataSet*: We selected a sample of 7 days between July 2005 and September 2005 uniformly from the gridMET dataset [6]. The files we chose were `vs_2005.nc` (Wind speed at 10 m) and `th_2005.nc` (Wind direction at 10 m) This specific time frame allows us to visualize both the impact of Hurricane Katrina, which struck the United States between August 23rd and September 1, 2005, and the changing seasonal patterns during this period. Observing these dates and interval captures both the immediate effects of the hurricane and the broader seasonal transitions.

2) *Implementation*: To transform and visualize the data, we used Matplotlib [8] and Cartopy [9] to visualize wind data through quiver plots on a map outline. The implementation includes directional arrows that represent wind direction at a 10-meter height, with wind speed magnitude indicated by color rather than varying arrow length, making the visualization consistent and clear. Additionally, we added coastlines and borders to provide geographical context, allowing the wind data to be accurately interpreted in relation to mapped locations.

3) *Experiments*:

   - Choosing Seasonal Periods: We selected three months July, August and September to observe seasonal variations in wind patterns. These months were chosen based on their tendency to exhibit significant shifts in wind speed and direction, which allowed for a comprehensive view of changing meteorological conditions.

   - Selecting the number of grid points: Once our study region was defined, we needed to resample grid points for plotting the quivers. Experimenting with different numbers (Fig 1, 2, 3), we found that a 30x40 point grid struck a balance between localized representation and vector visibility of meteorological conditions. This configuration allowed us to capture detailed regional patterns while ensuring that vectors were neither overly dense nor sparse.

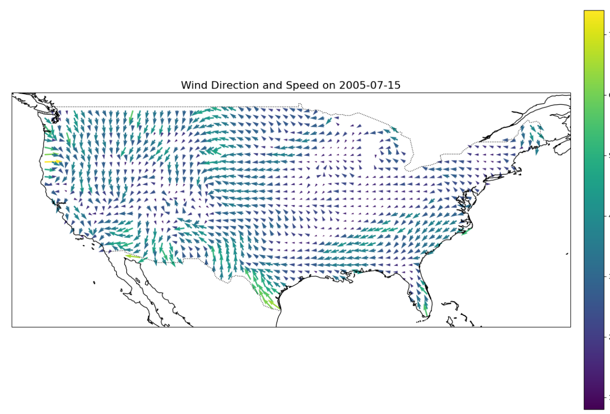

Fig 1. Resampling with a 20x20 grid. While vector directions are visible, it is difficult to discern wind magnitudes.

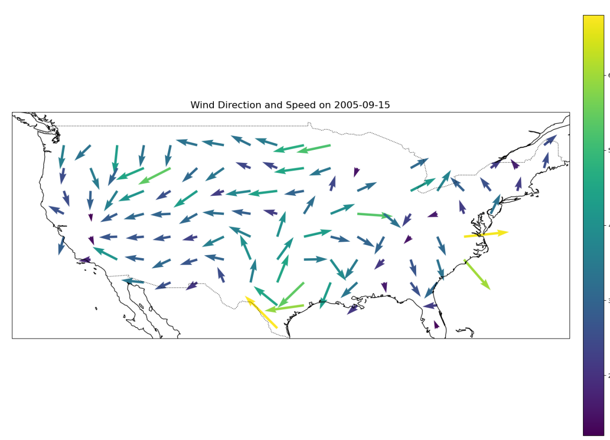

Fig 2. Resampling with a 60x70 grid. Although magnitude differences are visible, the sparse spacing of points limits meaningful interpretation.

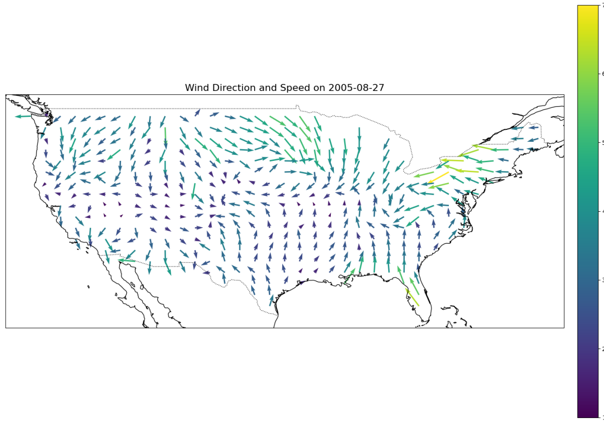

Fig 3. Resampling with a 30x40 grid. The vectors maintain their size for clear visibility.

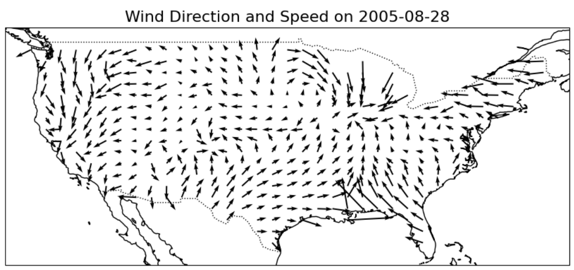

Fig 4. Quiver plot without color mapping, which lacks the clarity to distinguish varying wind speeds effectively.

- Defining Plot Frequency: To gain insights and avoid clutter, we selected specific days from July, August, and September: July 15, August 25 to September 1, and September 15. This selection provided a focused view, allowing us to capture patterns and shifts in wind dynamics around Hurricane Katrina's impact period and observe broader seasonal trends.

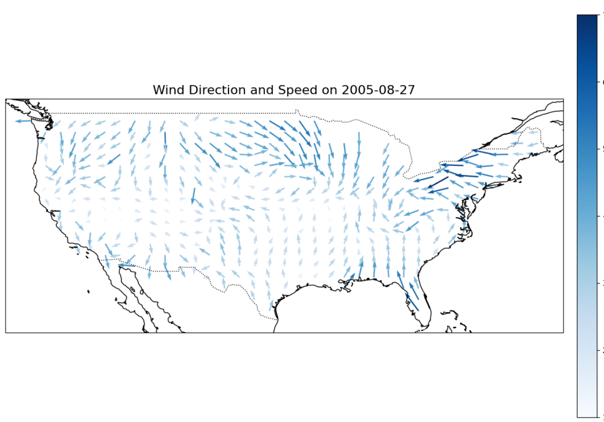

Fig 5. Quiver plot using the 'Blues' colormap instead of 'Viridis.' It is evident that this colormap is less suitable for effectively representing the wind speed variations in this plot.

- Resampling Grid Density: To maintain visual clarity, we resampled the wind data on a sparse grid, plotting every 30th latitude and 40th longitude point. After testing finer and coarser grids, we found this density best displayed regional patterns without sacrificing readability.

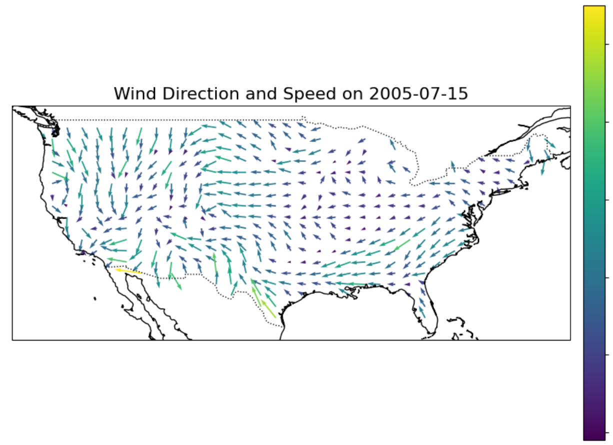

Fig 6. Plot generated for 15 July 2005

- Choosing Color Mapping: For representing wind magnitude, we experimented with various color mappings. Among the options, the perceptually uniform colormap 'viridis' (Fig 3.) was selected, as it provided a clear contrast across wind speed variations, effectively distinguishing areas of high and low magnitude, particularly in comparison to the 'Blues' colormap (Fig 4.). We also experimented with a quiver plot using plain black arrows without any color mapping. However, this approach proved less informative, as it failed to convey the variations in wind speed visually, making it difficult to identify patterns of intensity. Thus, 'viridis' was chosen to provide a more insightful and comprehensive representation of the data.

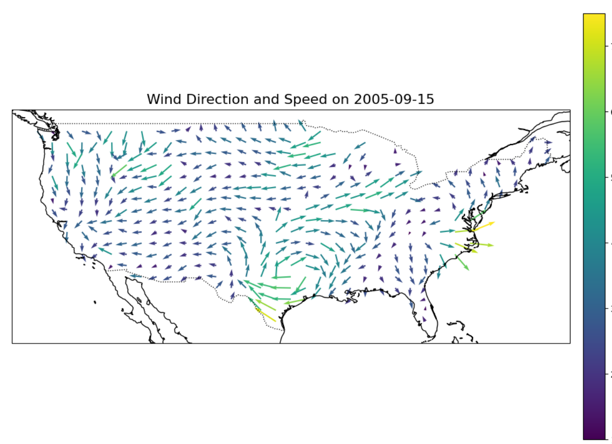

Fig 7. Plot generated for 15 September 2005

- Focusing on Hurricane Katrina: Recognizing the impact of Hurricane Katrina, which occurred between

August 25 (Fig 8.) and September 1 (Fig 14.), 2005, we incorporated specific analysis of this period. By visualizing wind patterns during these dates, we aimed to capture the storm's intensity and structure. This focused approach allowed us to highlight the unique meteorological conditions associated with Katrina.

4) *Generated Plots*: We created quiver plots to visualize wind patterns on selected days in July, August, and September 2005. The plots display wind direction through arrows, with colors representing the wind speed magnitude. Each plot is overlaid on a regional map to provide geographical context. Additionally, we included plots using the 'Blues' colormap and one with just black color applied to the arrows for comparison.

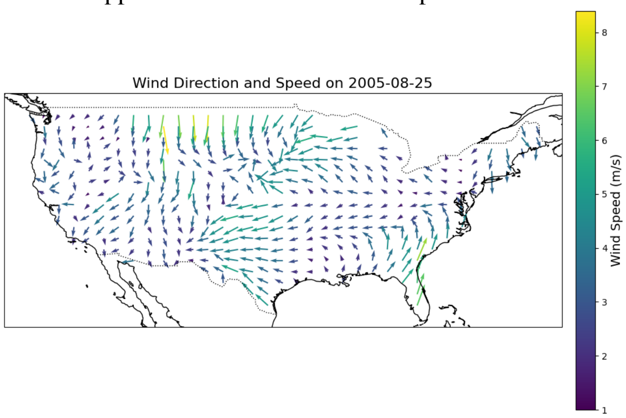

Fig 8. Plot generated for 25 August 2005

5) *Observations*: Our observations include:

- Wind speeds show significant increases, particularly on August 29 (Fig 11.), reflecting the intense impact of Hurricane Katrina with values exceeding 20 m/s near the Gulf of Mexico (approx. 29°N, 90°W). This peak day demonstrates the high winds associated with the storm's landfall on the northern Gulf Coast.

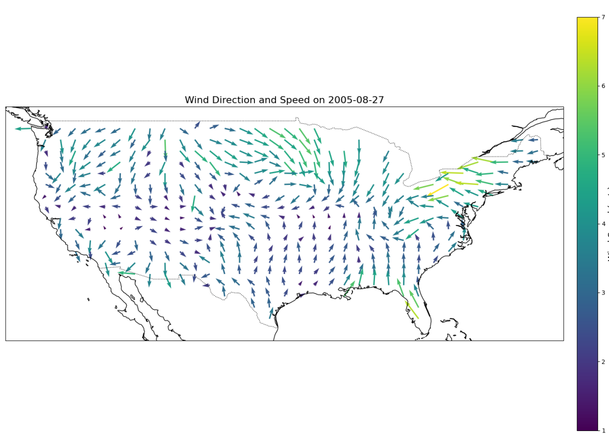

Fig 9. Plot generated for 27 August 2005

- The initial landfall of Hurricane Katrina occurred on August 25 along the southeast Florida coast near Miami (approx. 25°N, 80°W) as a Category 1 hurricane. This early landfall was characterized by moderate wind speeds and set the stage for the storm's re-intensification over the Gulf of Mexico.
- Wind directions predominantly exhibit a southwest flow around the time of landfall on August 29, particularly around Louisiana (30°N, 92°W) and Mississippi (32°N, 89°W). This counterclockwise flow typical of cyclones is clearly visible, underscoring the storm's large impact on regional wind patterns.

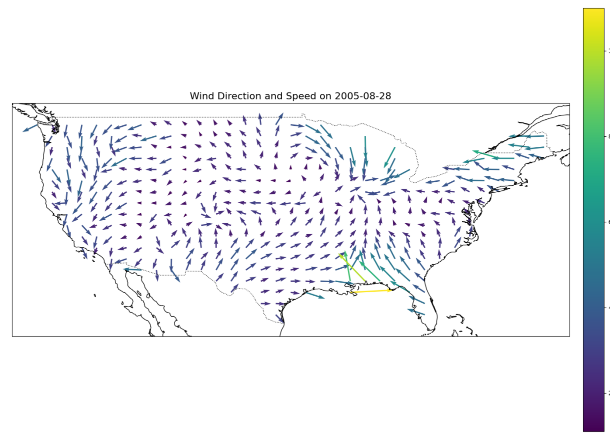

Fig 10. Plot generated for 28 August 2005

- On August 27 and 28 (Fig 9, 10), prior to its Gulf Coast landfall, there is a noticeable rise in wind speeds across the Gulf region, signaling the storm's approach. Rising winds in this area (centered around 25°N, 87°W) reflect the storm's intensification over warm Gulf waters as it gained strength.

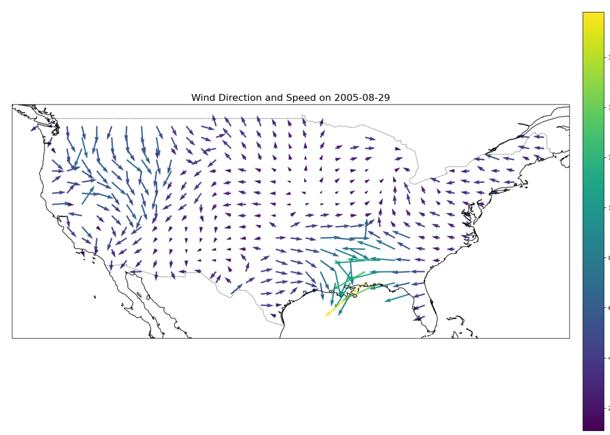

Fig 11. Plot generated for 29 August 2005

- Following the peak winds on August 29, there is a marked reduction in wind speeds in early September. This reduction signifies the dissipation of Katrina's energy and the return to more stable conditions as the storm weakened over land, eventually dissipating over central Mississippi.
- The intense winds and shifting directions on August 29 highlight the storm's widespread impact. Katrina's

trajectory from its initial landfall in Florida to its devastating impact on the Gulf Coast (29°N, 90°W to 33°N, 88°W) exemplifies its strength and reach across multiple states.

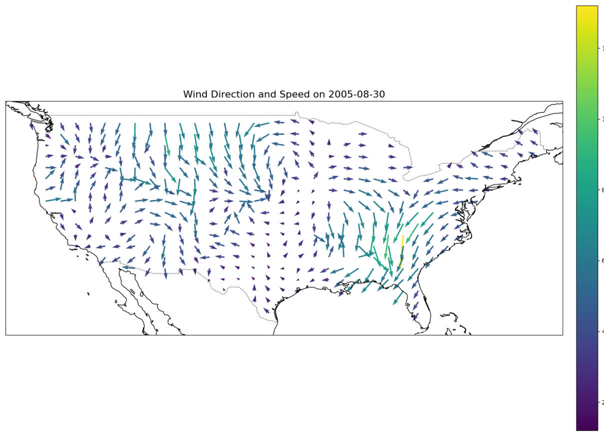

Fig 12. Plot generated for 30 August 2005

- Observations on July 15 (Fig 6.), August 25-31 (Fig 8-14.), and September 15 (Fig 7.) also highlight seasonal transitions. More stable, lower wind speeds in July gradually intensify through August, reflecting typical late-summer patterns that culminate with heightened storm activity. By September 15, wind patterns show further stabilization as the season transitions toward fall, moving away from the peak hurricane period.

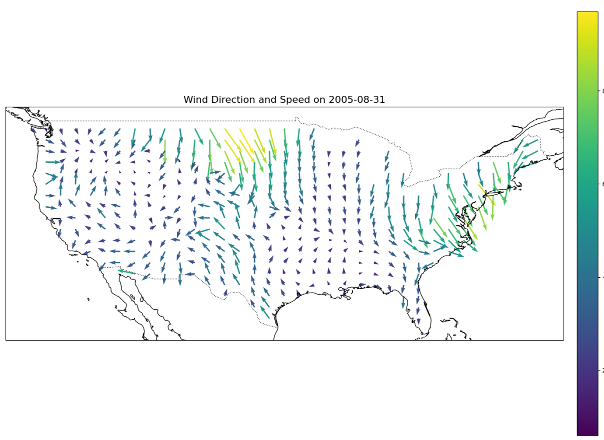

Fig 13. Plot generated for 31 August 2005

- The variations in wind speed and direction over these months provide insight into broader seasonal shifts. The significant activity in August, contrasted with relatively calmer conditions in July and September, highlights the cyclical nature of wind patterns during the late summer hurricane season in the Gulf region.

In summary, the observations highlight the dramatic wind patterns and impacts of Hurricane Katrina, particularly during its landfall in late August 2005. The data showcases the storm's intensity during both its initial landfall in Florida and its Gulf Coast landfall, the characteristic counterclockwise circulation, and subsequent dissipation, as well as the broader seasonal shifts typical of this period in the Gulf region.

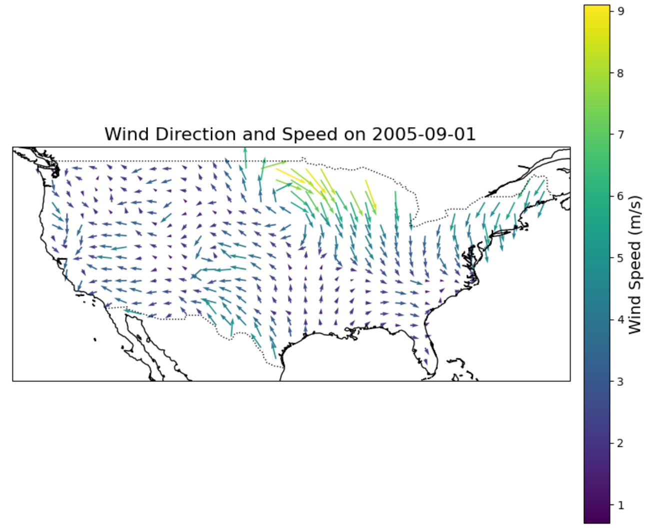

Fig 14. Plot generated for 1 September 2005

B. Colormap

Colormap assigns colors to data values in a visualization, highlighting patterns across a range. By mapping data points to specific colors, it visually distinguishes features.

Dataset : We have used the gridMET dataset [6] which provides daily meteorological data for a period of over 40 years. We have sampled 7 days over the period of August and September 2005.
The days sampled are:
• 25th August • 27th August • 28th August • 29th August
• 30th August • 31st August • 1st September
This specific time frame was chosen to visualize the impact of "Hurricane Katrina" [7], which struck the Gulf Coast of United States between August 25th and August 31, 2005, and the changing seasonal patterns during this period.

Data Preprocessing: The files `pr_2005.nc` and `vs_2005.nc` (precipitation amount in mm and wind speed in m/s, respectively) were downloaded and preprocessed using the netCDF4 library in Python. The relevant variables for the sampled days were then extracted.We also set specific latitude and longitude ticks with labels for northern latitudes and western longitudes on the plot.

Implementation: To analyze the hurricane, we felt that *precipitation amount* and *wind speed* would be the most relevant variables. We then plotted discrete,continuous and logarithmic colormaps and experimented with different kinds of divergent and sequential colormaps for the sampled dates. Matplotlib's pcolormesh was used to plot.

For each sampled day, we calculated the maximum and minimum values of these two variables, then determined the global minimum and maximum across all the sampled days. This would let us analyze the hurricane in a better way.

Trying out different Color Mapping Scales:

- **Continuous**
  We have used continuous color scale here in Fig.15 for the precipitation amount variable. A smooth progression of values over time is observed, where each color corresponds to a specific range of values in the data.

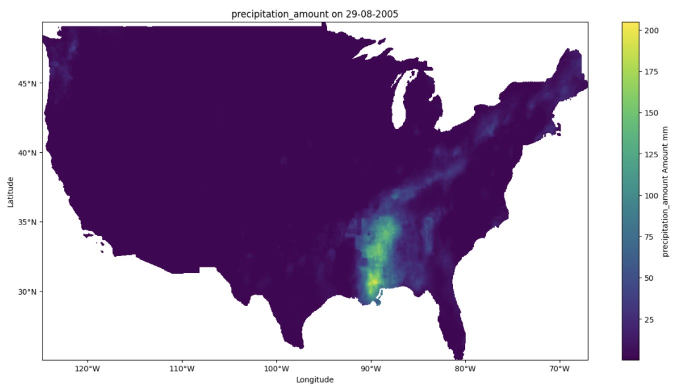

Fig. 15. Continuous Color scale

- **Discrete**
  We have used discrete color scale here in Fig.16 for the precipitation amount variable ,where specific values are represented by distinct colors. The highest values in the data are marked in red.

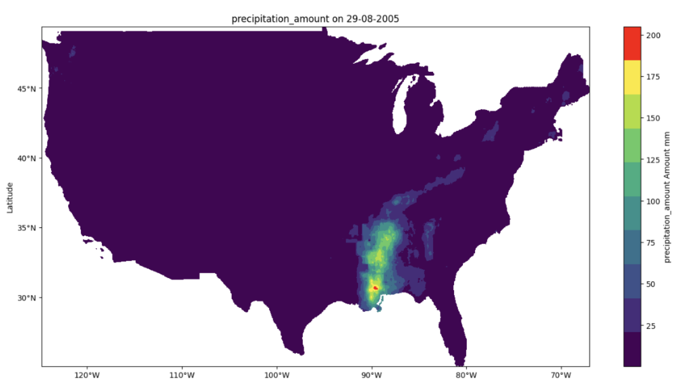

Fig. 16. Discrete Color Scale

- **Logarithmic**
  We have used Logarithmic color scale here in Fig.17 for the precipitation amount variable.This scale is useful for compressing large variations in data, especially when there are extreme high or low values.Since our data does not have such extreme values, using a logarithmic scale doesn't provide much advantage in this case.

Hence, the continuous color scale is the most effective for visualizing the precipitation and wind speed data. It allowed

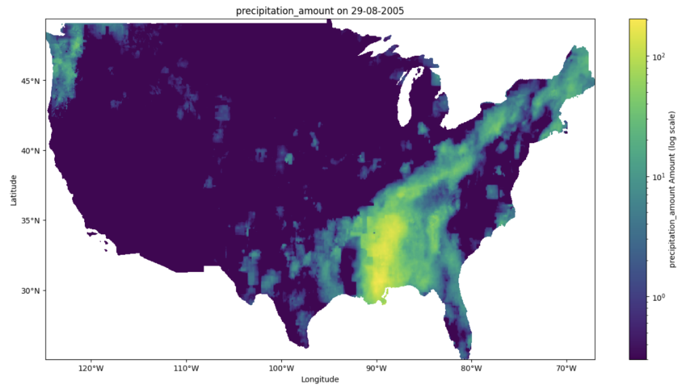

Fig. 17. Logarithmic Color scale

for a smooth and clear representation of the variations in both variables, making it easy to identify patterns and trends

Trying out different divergent and sequential colormaps:

- **Divergent colormaps**
  We experimented with several divergent colormaps for the Wind Speed variable, including bwr Fig.18, coolwarm, seismic and BrBG Fig.19. A few of them are shown here.
- **Sequential colormaps**
  We experimented with several Sequential colormaps for the Wind Speed variable, including inferno Fig.21, cividis, viridis, blues Fig.20 .A few of them are shown here.

Diverging colormaps are typically used when the data has a clear midpoint, highlighting differences from this central point. Sequential colormaps are used when data values progress continuously from low to high or high to low, without a clear midpoint. Since our dataset does not have a defined midpoint and focuses on continuous variations , we chose to use sequential colormaps.

We Proceed with Viridis Sequential Colormap with Continuous scale.

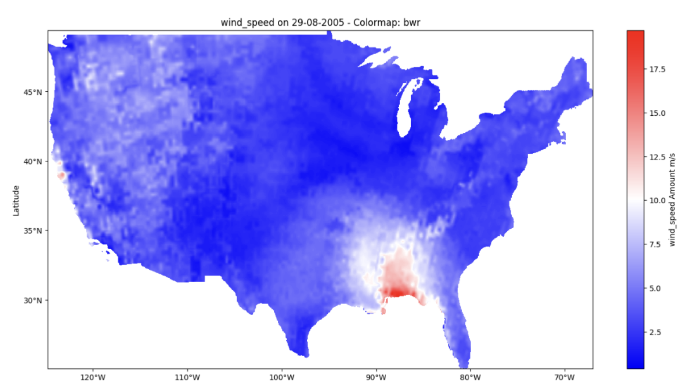

Fig. 18. Bwr-Divergent Colormap

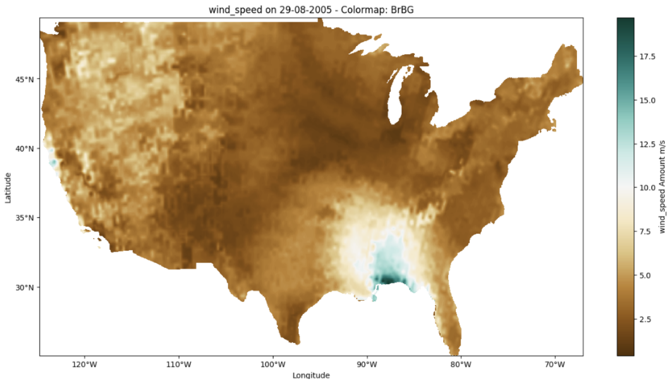

Fig. 19. BrBg-Divergent Colormap

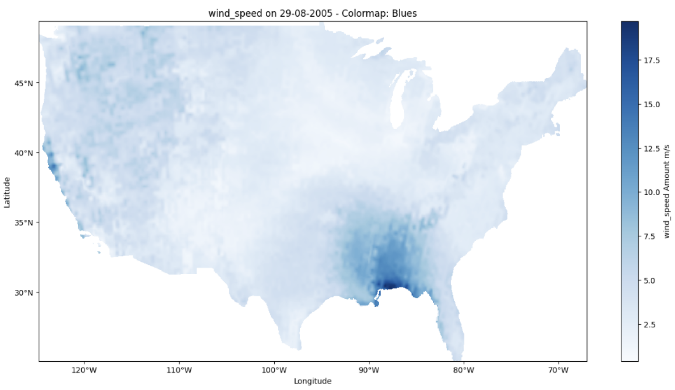

Fig. 20. Blues-Sequential Colormap

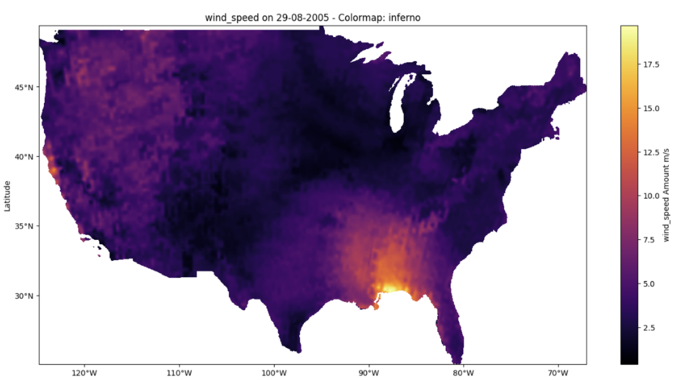

Fig. 21. Inferno-Sequential Colormap

Analyzing the sampled dates:

- Katrina Storm made [7] its initial landfall along the southeast Florida coast on "August 25th" as a Category 1 storm. As shown clearly in the Fig.22, the wind speed is slightly high along the southeast Florida coast at approximately **Lat 27°N** and **Lon 80°W** , which is clearly highlighted by light yellow color in fig .
- After moving west across southern Florida and into the warm waters of the Gulf of Mexico, Hurricane Katrina rapidly intensified, reaching Category 5 status on "August 28th". This intensification is clearly represented in the Fig.23 and Fig.24 at approximately **Lat 30°N** and **Lon 90°W**, where both wind speed and precipitation levels peak, shown in dark yellow around the Gulf of Mexico. The continuous color scale in this plot provides a precise representation of the varying intensities of precipitation, with darker yellow tones highlighting areas of extreme rainfall and high wind speed.
The use of a smooth color transition allows for easy identification of regions most affected by the storm.

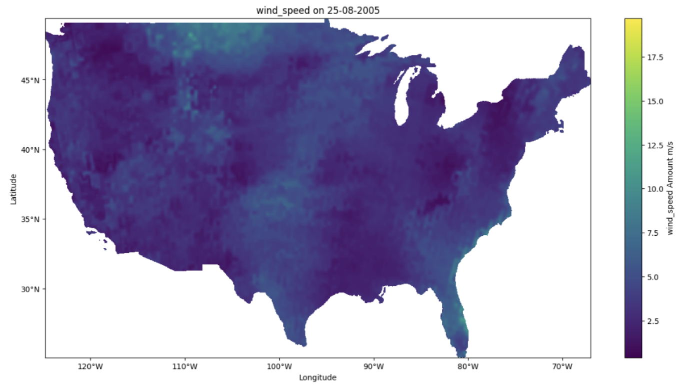

Fig. 22. Wind Speed on 25th Aug, 2005

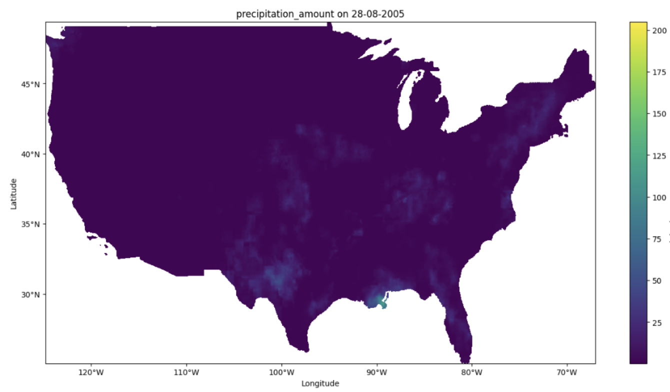

Fig. 23. Precipitation Amount on 28th Aug, 2005

- The storm continued north-northeast and moved inland into southern Mississippi, around **Lat 32°N** and **Lon 92°W**, on "August 29th". The impact of the hurricane is visible in the Fig.25 and Fig.26, where both wind speed and precipitation are represented in dark yellow, indicating the storm's severity as it moved further inland. This color intensity on the continuous scale highlights the extreme conditions in these areas, where the darkest yellow regions mark the highest wind speeds and heaviest rainfall. The chosen colormap provides a clear gradient to emphasize the storm's intensity, with progressively

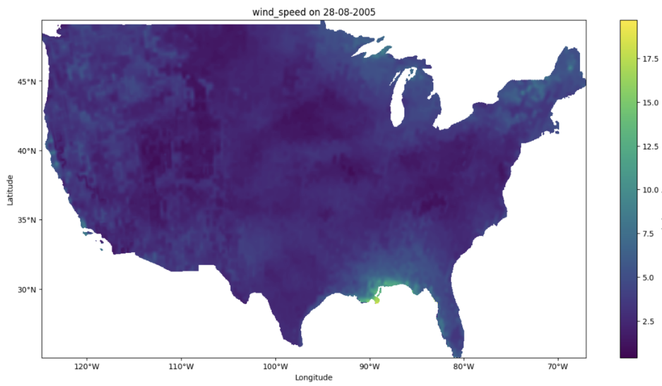

Fig. 24. Wind Speed on 28th Aug, 2005

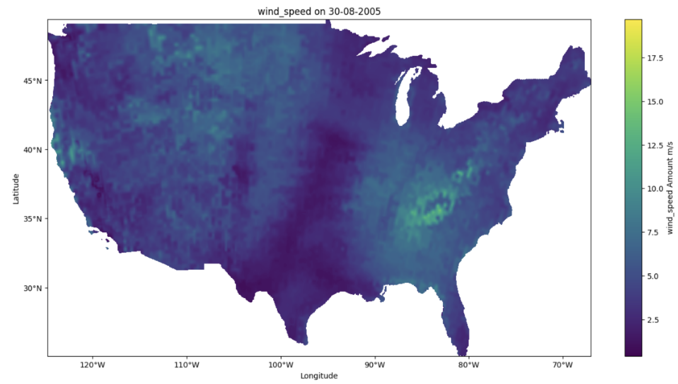

Fig. 27. Wind Speed on 30th Aug, 2005

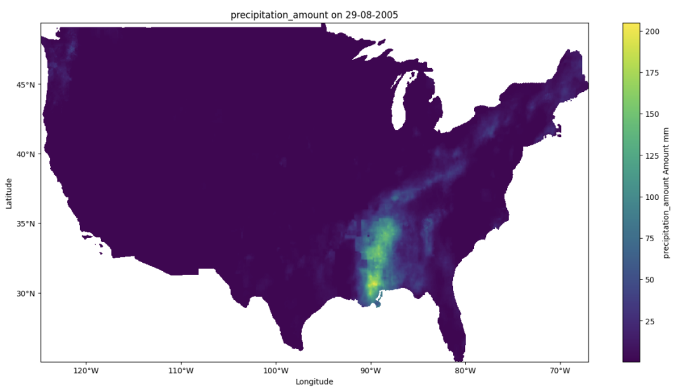

Fig. 25. Precipitation Amount on 29th Aug, 2005

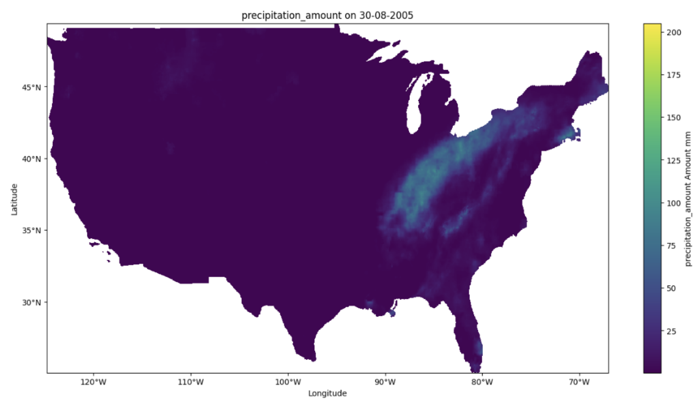

Fig. 28. Precipitation Amount on 30th Aug, 2005

darker tones capturing the devastating force of Hurricane Katrina.

- On "August 30", Hurricane Katrina slightly weakened a little as it moved inland to the east, transitioning from a hurricane to a tropical depression. As shown in the Fig.28, the area from **Lon 80°W to 90°W** and **Lat 35°N to 40°N** is marked by yellow, indicating the intense precipitation Katrina brought to regions like the Ohio Valley and Mid-Atlantic, which led to localized flooding and widespread disruptions.

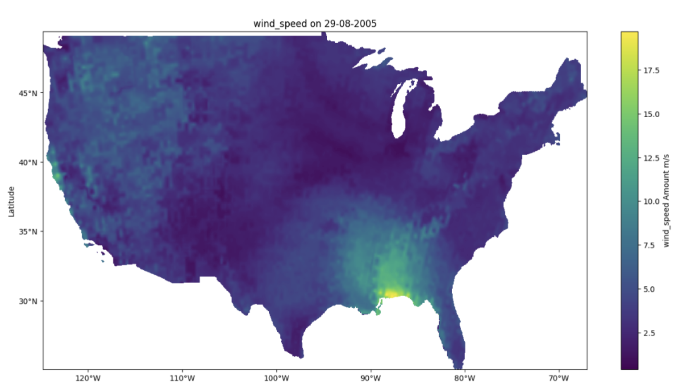

Fig. 26. Wind Speed on 29th Aug, 2005

In Fig.27 the wind speed's color gradient clearly shows how Hurricane Katrina's impact decreased, but still remained strong. The darker yellow areas highlight the heavy rainfall, showing how far the storm's effects reached. The gradient helps us see where and how intense the windspeed was, letting us follow the path and size of Katrina's influence, even as it weakened.

- On "August 31", as the Hurricane Katrina moved northeastward, they were absorbed by an approaching cold front near the Great Lakes region, **Lon 75°W Lat 45°N** . In the Fig.29, light precipitation associated with the weakening storm as seen in Fig.30 system can be observed, depicted in light yellow across the area. This indicates lingering moisture from Katrina, even as it dissipated, contributing to rainfall over the region. The interaction with the cold front enhanced precipitation in some localized areas, as shown by yellow shading in the plot.
  The Hurricane got over on 31st of August.

- On "September 1" , Fig.31 show no signs of wind or rain, which indicates that Hurricane Katrina had completely ended. In both plots, there are none of the

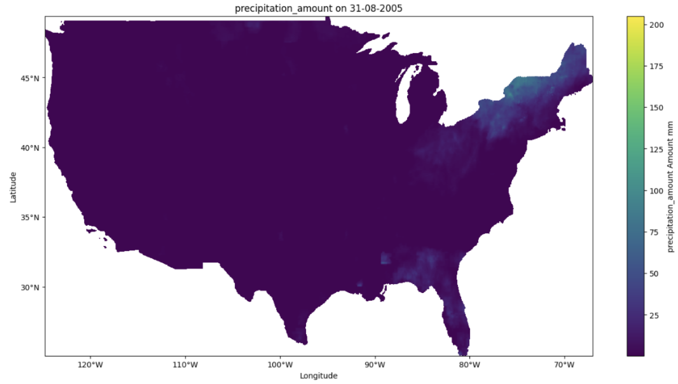

Fig. 29. Precipitation Amount on 31st Aug, 2005

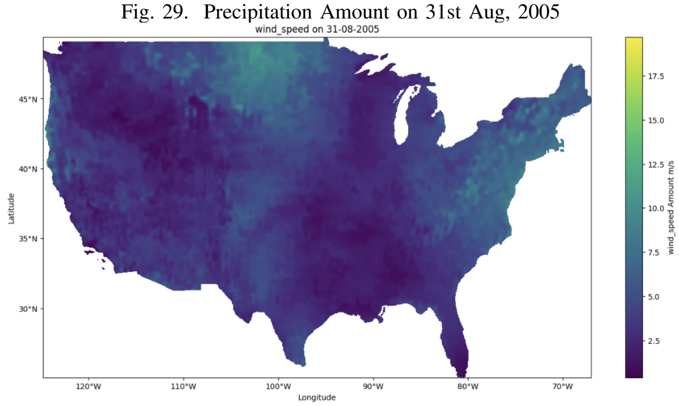

Fig. 30. Wind Speed on 31st Aug, 2005

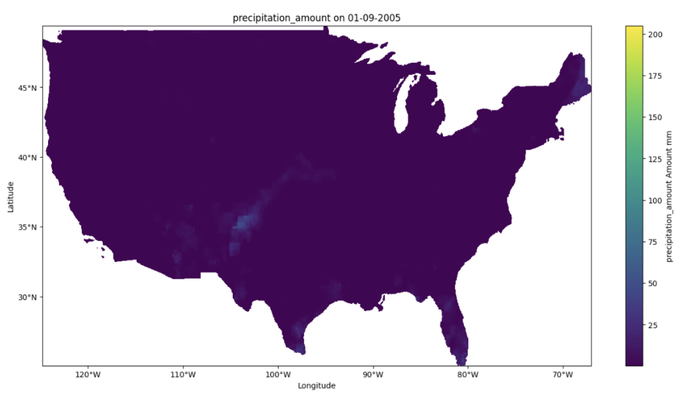

Fig. 31. Precipitation Amount on 1st Sept, 2005

typical features of a storm—no strong winds or heavy rain bands that were visible in earlier days. The calm and clear conditions shown in the images confirm that Katrina had fully dissipated after being absorbed by a cold front near the Great Lakes on August 31. These plots highlight the quiet weather left behind after the storm

This is our analysis and observations on Hurricane Katrina using colormap, which first struck on August 25 and continued until August 31. The chosen color scale and colormap plots clearly show its movement and impact over these days, capturing the storm's path and intensity as it traveled, before finally dissipating.

Animation GIFs:
We have made animation GIFs for the sampled days, showing the variations in precipitation and wind speed variables across the days. We have submitted it in our folder.

C. Contour Map

A contour map represents data values as lines or filled regions, allowing viewers to see gradients and variations across a range. By mapping data points to specific contour levels, it effectively highlights differences in value and reveals patterns across the visualized area.

Dataset : We have used the gridMET dataset which provides daily meteorological data for a period of over 40 years. We have sampled 7 days over the period of August and September 2005.

The days sampled are:
• 25th August • 27th August • 28th August • 29th August
• 30th August • 31st August • 1st September

This specific time frame was chosen to visualize the impact of Hurricane Katrina, which struck the Gulf Coast of United States between August 25th and August 31, 2005, and the changing seasonal patterns during this period.

Data Preprocessing: The files `pr_2005.nc` and `vs_2005.nc` (precipitation amount in mm and wind speed in m/s, respectively) were downloaded and preprocessed using the netCDF4 library in Python. The relevant variables for the sampled days were then extracted.

Implementation : To analyze the hurricane, we selected precipitation amount and wind speed as the most relevant variables. We first created contour maps displaying only contour lines using the Marching Squares algorithm, followed by filled contour maps using contourf to better illustrate the variable distribution. We experimented with various colormap scales (divergent and sequential) for each sampled date to capture patterns effectively. For each day in our sample, we calculated the minimum and maximum values of these variables, then identified the global minimum and maximum across all dates to ensure consistent analysis. For plotting, we used matplotlib.pyplot and cartopy libraries.

In our contour plotting efforts, we experimented with two methods: contour fill and marching squares.

- **Marching Squares:** We generated contour plots (Fig. 32, Fig. 33, Fig. 34, Fig. 35, & Fig. 36) employing the marching squares algorithm and conducted experiments with different plot configurations, including colormap, contour levels, and line widths to achieve the most visually appealing plot. However, we observed that these visualizations posed challenges for regular readers to

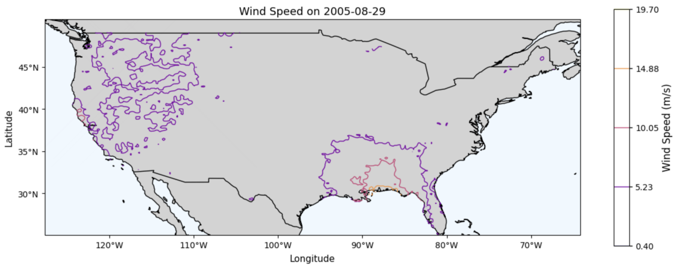

Fig. 32. Contour plot showing wind speed on August 29th, generated with 5 levels. The plot uses a line width of 1 and applies the ‘plasma’ colormap to represent variations in wind speed.

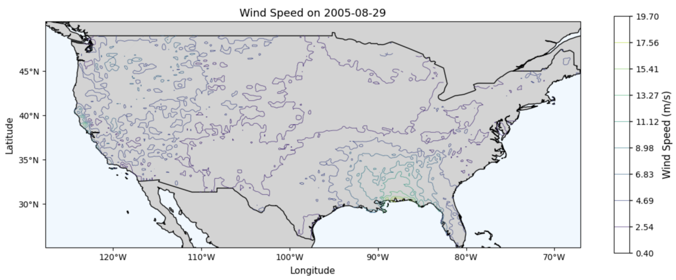

Fig. 33. Contour plot showing wind speed on August 29th, generated with 10 levels. The plot uses a line width of 0.5 and applies the ‘viridis’ colormap to represent variations in wind speed.

comprehend when utilizing the marching squares technique.

- **Contour Fill**: We generated contour plots using the contour fill method, and an example is shown in Fig. 37. Various settings, such as colormap and contour levels, were adjusted to optimize the visualization. After experimentation, we found the optimal plot was achieved using the `'viridis'` colormap with 10 contour levels.
  - *Choice of Colormap:* A divergent colormap was not suitable because there is no central reference point in the data. Therefore, we used sequential colormaps and experimented with options like `jet`, `coolwarm`, `plasma` (Fig. 38), and `viridis` (Fig. 37). While `viridis` and `plasma` produced similar results, we chose `viridis` due to its enhanced perceptual uniformity and clarity.

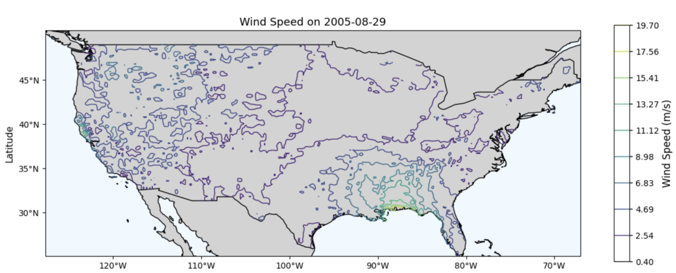

Fig. 34. Contour plot showing wind speed on August 29th, generated with 10 levels. The plot uses a line width of 1 and applies the ‘viridis’ colormap to represent variations in wind speed.

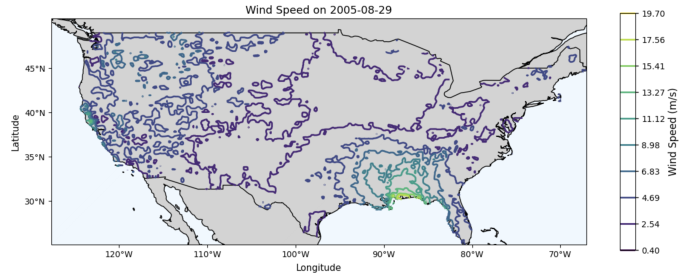

Fig. 35. Contour plot showing wind speed on August 29th, generated with 10 levels. The plot uses a line width of 2 and applies the ‘viridis’ colormap to represent variations in wind speed.

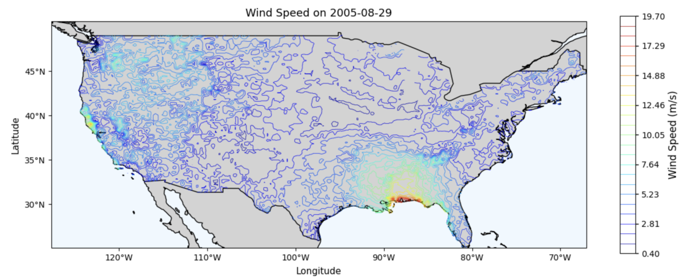

Fig. 36. Contour plot showing wind speed on August 29th, generated with 25 levels. The plot uses a line width of 0.5 and applies the ‘jet’ colormap to represent variations in wind speed.

  - *Choice of Number of Levels:* We experimented with 5 (Fig. 39),10 (Fig. 38) and more contour levels. Using 5 levels may ignore some important trends in the data. Upon further experimentation, we found that using more than 10 levels made the graph less visually appealing and was not necessary to effectively represent the phenomenon.

  For the plot, scan line fill is used as the exact regions do not matter in big distances like in a map.

Analyzing the Sampled Dates with Contour Fill Maps :

- August 25 - Initial Landfall : Hurricane Katrina first made landfall along the southeast Florida coast on August 25 as a Category 1 storm. In the contour fill map (Fig. 40 & Fig. 41), the storm’s intensity is represented by distinct color

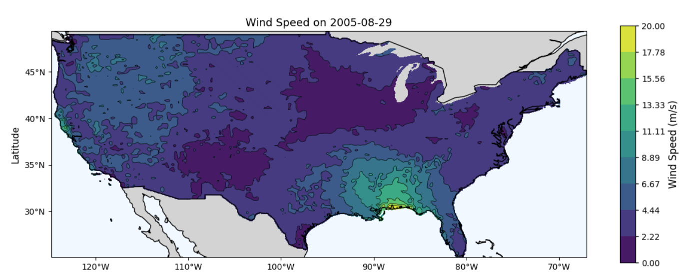

Fig. 37. Contour plot showing wind speed on August 29th, generated with 10 levels. The plot applies the ‘viridis’ colormap to represent variations in wind speed.

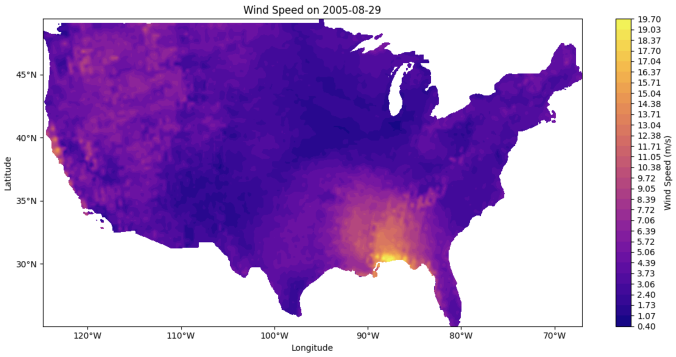

Fig. 38. Contour plot showing wind speed on August 29th, generated with 10 levels. The plot applies the 'plasma' colormap to represent variations in wind speed.

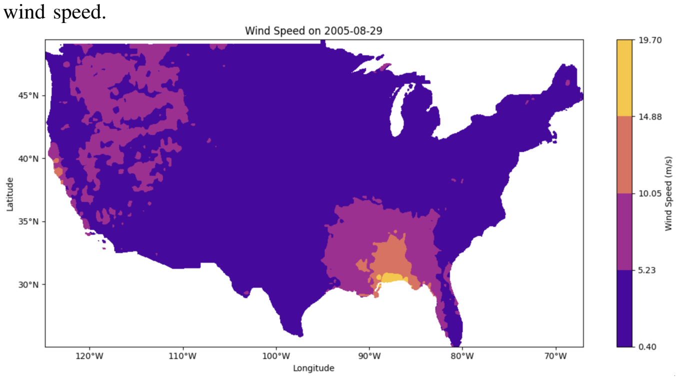

Fig. 39. Contour plot showing wind speed on August 29th, generated with 5 levels. The plot applies the 'plasma' colormap to represent variations in wind speed.

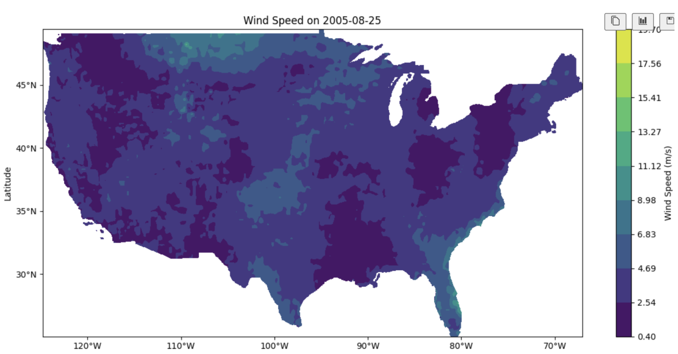

Fig. 40. Wind Speed on August 25th

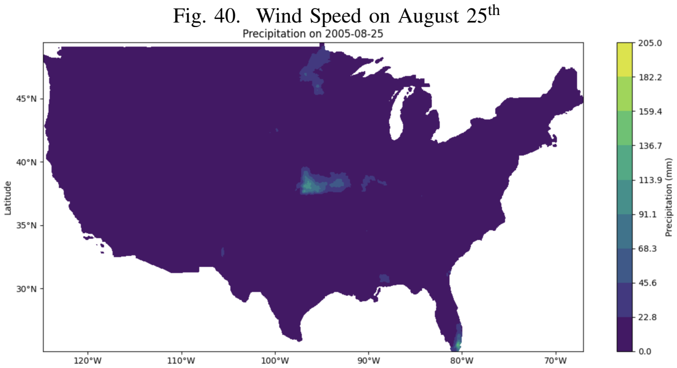

Fig. 41. Precipitation on August 25th

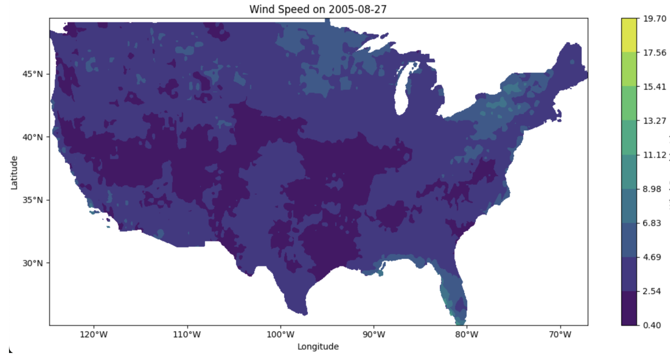

Fig. 42. Wind Speed on August 27th

bands that highlight higher wind speeds near the southeast Florida coast at approximately Lat 27°N and Lon 80°W. This area displays darker color fills, indicating elevated wind and rainfall in this region as the storm began.

- August 27 - Early Wind Patterns : The map in Fig.42 shows wind speeds across the United States on August 27, 2005, with darker shades indicating lower speeds and lighter shades indicating higher speeds. Stronger winds, around 9-12 m/s, are visible near the southeastern coast, especially around Florida and the Gulf of Mexico, as Hurricane Katrina started to strengthen. This contour fill map highlights early impact zones and hints at the storm's impending intensification.
- August 28 - Intensification Over the Gulf : As Katrina moved across southern Florida and into the warm waters of the Gulf of Mexico, it rapidly intensified, reaching Category 5 status on August 28. This intensification is evident in contour fill maps (Fig.43 & Fig.44), where color shades like cyan, greenish, yellowish fill—indicating higher values—are concentrated around Lat 30°N and Lon 90°W. The intense color fills in this region represent the peak levels of wind speed and precipitation, particularly around the Gulf of Mexico. The contour fill enables us to see the localized regions of extreme impact and the storm's strengthening as it approached land.

- August 29 - Landfall in Mississippi : On August 29, Hurricane Katrina made landfall in southern Mississippi, around Lat 32°N and Lon 92°W. In the contour fill maps (Fig. 45 & Fig.46), high-intensity areas are depicted by deep, saturated colors, marking the highest levels of wind speed and precipitation in these regions. The appearence of lighter shades of green and yellow illustrates the severity of the storm's impact as it moved further inland. This color gradient effectively highlights regions of extreme conditions, with the deepest hues corresponding to the strongest winds and heaviest rainfall as the storm progressed.

.

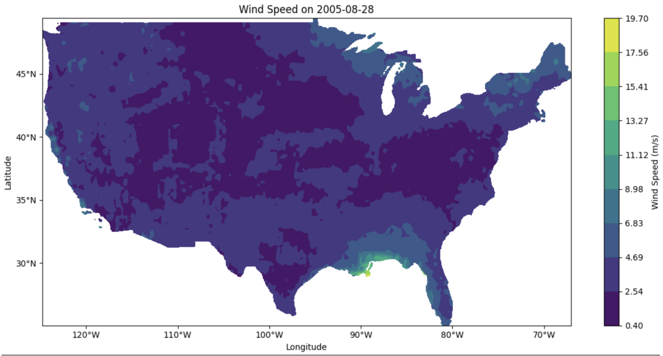

Fig. 43. Wind Speed on August 28th

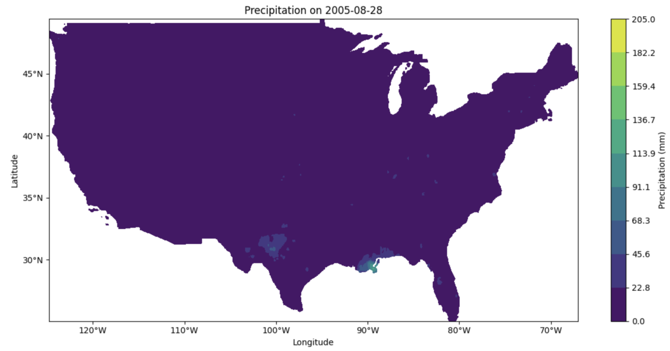

Fig. 44. Precipitation on August 28th

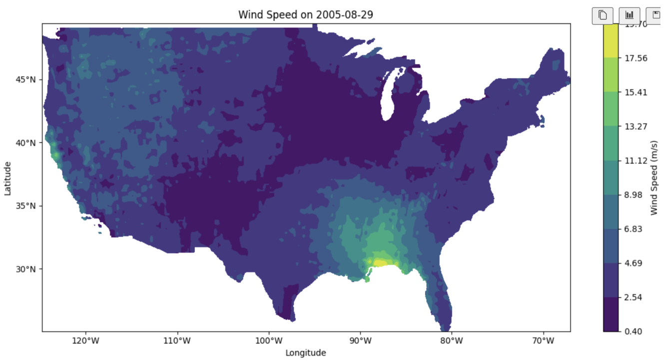

Fig. 45. Wind Speed on August 29th

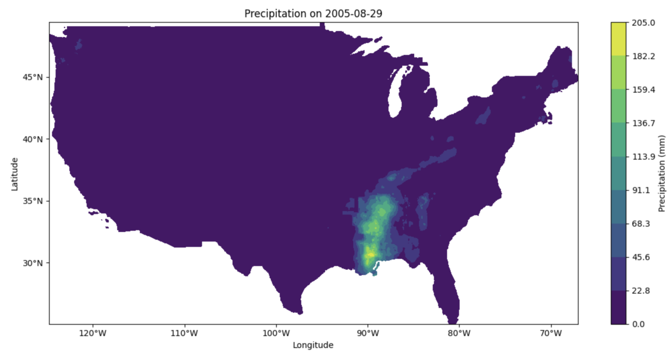

Fig. 46. Precipitation on August 29th

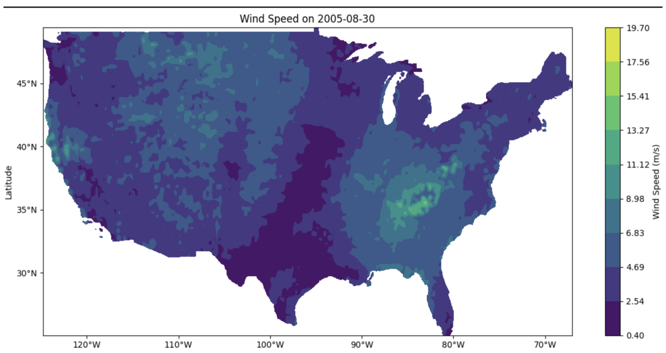

Fig. 47. Wind Speed on August 30th

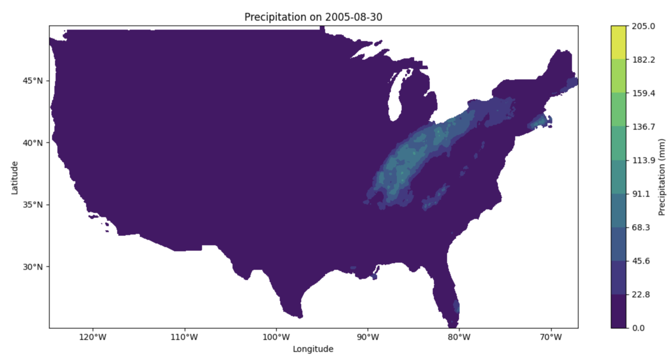

Fig. 48. Precipitation on August 30th

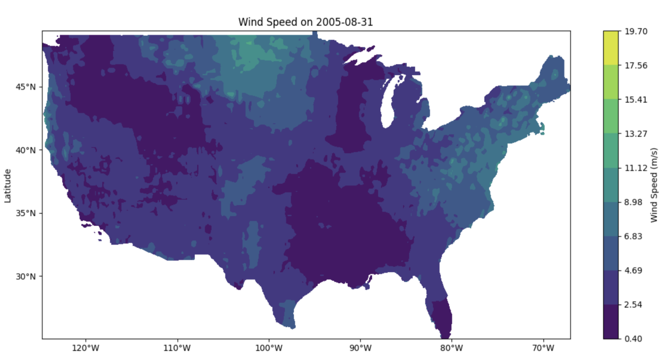

Fig. 49. Wind Speed on August 31th

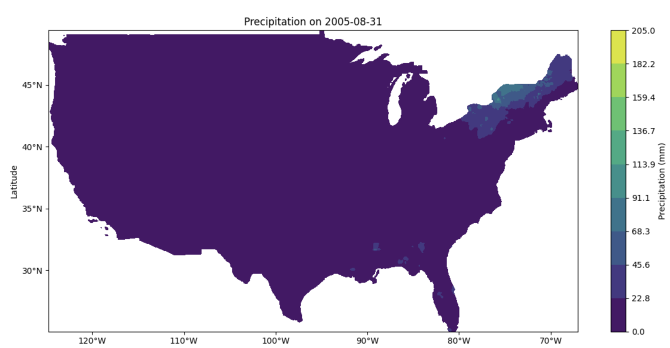

Fig. 50. Precipitation on August 31th

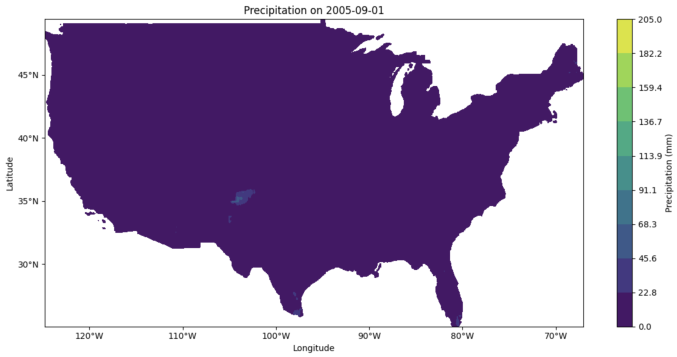

Fig. 51. Precipitation on September 1th

- August 30 - Moving Inland as a Tropical Depression : On August 30, Katrina weakened as it continued moving inland, transitioning into a tropical depression. In Fig.48, the contour fill map shows regions of broad spread of intense precipitation from Lon 80°W to 90°W and Lat 35°N to 40°N, impacting regions like the Ohio Valley and Mid-Atlantic. While the color fills remain lighter, indicating substantial rainfall and wind speed, they are less concentrated than in earlier figures, suggesting a decrease in storm intensity. Fig.47 reveals the contour fill for wind speed, which now displays a wider distribution with less saturation, showing the diminished but still extensive wind effects as Katrina moved northward.

- August 31 - Dissipation Near the Great Lakes : By August 31, Katrina had significantly weakened and was moving northeastward, where it was eventually absorbed by a cold front near the Great Lakes region around Lon 75°W and Lat 45°N. In Fig.49 and Fig.50, the contour fill maps show dareker shades representing residual, light precipitation and comparitively lower wind speeds across this area. This pattern, marked by lighter color fills, indicates lingering moisture as the storm dissipated, with additional rainfall in some localized areas due to interaction with the cold front.

- September 1 - Complete Dissipation : On September 1, Fig. 51.
  shows a lack of bright, lighter contour fills in the Viridis colormap, confirming that Hurricane Katrina had fully dissipated. In contrast to the brighter yellow and green fills seen in previous days, which indicated high wind speeds, the darker purple shades in this figure show low values, representing calm atmospheric conditions with minimal wind activity.

In conclusion, the contour fill maps, particularly with the Viridis colormap, provided a powerful tool for understanding the full life cycle of Hurricane Katrina across its active dates. By displaying data across latitude and longitude coordinates, these maps visually captured the hurricane's path and intensity over time—from initial landfall to peak strength and eventual dissipation. The Viridis colormap, with its gradient from dark purple for lower values to bright yellow for higher intensities, allowed us to clearly identify regions with high wind speeds and heavy rainfall at different stages of the storm. Contour fills smoothed out the data into continuous color bands, making it easy to see patterns of intensity changes and how they were distributed spatially. This approach was essential for tracking Katrina's progression and influence over vast areas, highlighting areas most affected and allowing us to visually follow its impact day by day.

Animation GIFs : We have made animation GIFs for the sampled days, showing the variations in precipitation and wind speed variables across the days. We have submitted it in our folder.

II. Information Visualization

A. Node Link Diagram

Dataset: Our dataset is the "Amazon Product Co-purchasing Network and Ground-truth Communities" [5]. It contains set of undirected edges between nodes. It is based on the "Customers Who Bought This Item Also Bought" feature of the Amazon website. If product i is frequently co-purchased with product j, the graph contains an undirected edge from i to j. The dataset consists of 334,863 nodes and 925,872 edges between them. It is an undirected graph , without any labels.

Preprocessing of Dataset: Since our dataset contains a large number of nodes, visualizing the entire dataset at once was not feasible. Therefore, we filtered the dataset to focus on the most relevant relationships. We used NetworkX to create an undirected graph from the data and identified the most popular product by calculating node degrees, which highlights products with the most connections. We then filtered the data to include direct (1st-level) and indirect (2nd-level) co-purchases of this popular product, capturing key relationships without unnecessary complexity. This preprocessing step reduced the dataset to 1,315 nodes and 1,813 edges, making it manageable for analysis and revealing significant purchasing patterns.

Implementation: We visualized the filtered dataset using different layout algorithms such as Force Atlas, Yifan Hu, Fruchterman-Reingold, Radial axis, OpenOrd using Gephi and experimented with different statistical methods in Gephi [4].Explanation regarding the same is mentioned below.

Layout Algorithms:

Force Atlas
This iterative algorithm works in the following way [1]:

1) Identify the nodes with the highest number of connections (the hubs) and push them apart from each other;
2) Pull the nodes that are connected to those hubs towards each other.

As a result, after multiple iterations, the most influential nodes are spread in the representation space while the smaller nodes that are connected to them are pulled towards those hubs, forming communities.
Gephi's Force Atlas algorithm parameters are Inertia, Repulsion strength, Attraction strength, Maximum Displacement, Gravity, Size, Speed.

The output of this algorithm can be seen in Fig.52.

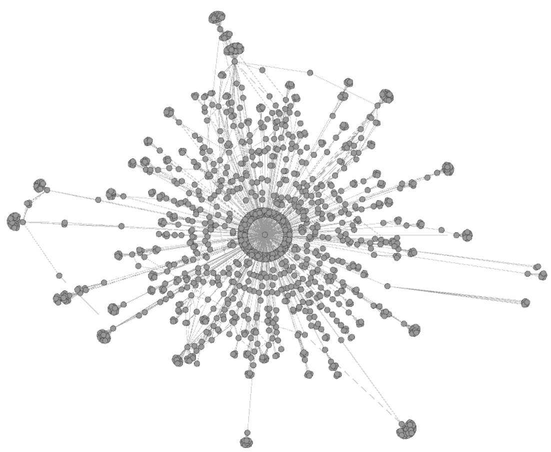

Fig. 52. Output of Force Atlas Algorithm

In the Force Atlas layout, the most popular product, being the most connected in the network, naturally appears at the center of the plot, with other co-purchased products arranged around it. This layout highlights the relationships, with closer products being more frequently co-purchased, providing a clear visualization of the network's structure based on the filtered dataset. The centrality of the most popular product is evident, displaying its significance in the co-purchase network. But , we can clearly see that the edges are cluttered in this layout and are not that clear.

In Fig.53., we applied both color and size rankings to nodes based on degree using Gephi's ranking options. Nodes with higher degrees are darker and larger, highlighting the most popular product at the center as the darkest and largest node. There are smaller clusters around it also , with dark nodes as local centers, making it easy to see important hubs within the co-purchase network.Since the most popular product has a much higher degree than the other nodes, the remaining nodes appear relatively smaller in comparison.

In Fig.54, we used Gephi's modularity feature to detect communities within the network. This groups products into clusters based on co-purchasing patterns, where products within the same community are more densely connected to each other than to products in other communities. Each community is color-coded, making it clear which groups of

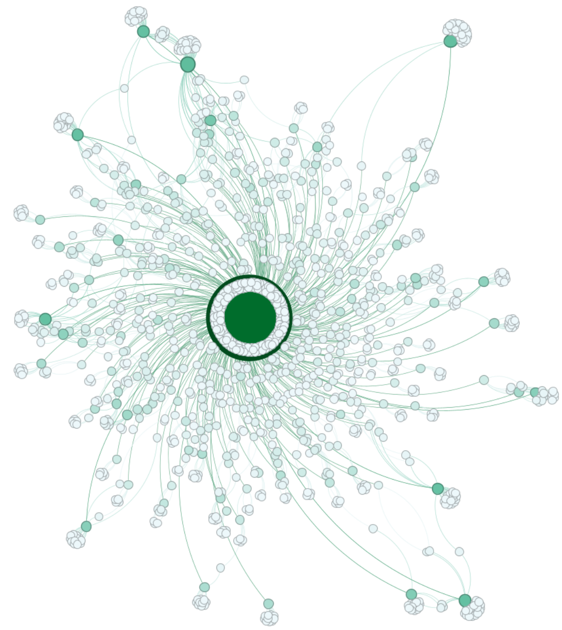

Fig. 53. Mapping the degree of each node to a color and Size

products are commonly bought together.

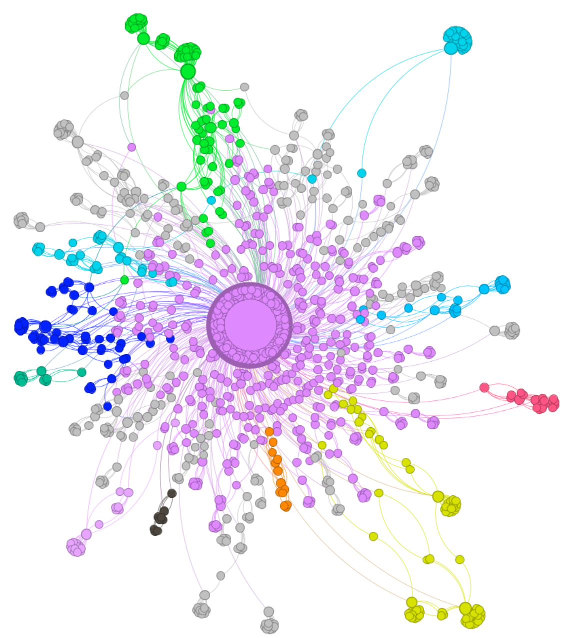

Fig. 54. Identifying clusters through modularity in the Force Atlas algorithm.

Fruchterman-Reingold

The Fruchterman-Reingold algorithm [2] is a force-directed graph drawing algorithm that positions the nodes of a graph in a way that visually reflects its structure by simulating physical forces. It treats edges as springs that pull connected nodes together while simultaneously considering repulsive forces between all pairs of nodes, which is visually appealing. Gephi's Fruchterman-Reingold algorithm parameters are Area, Gravity and Speed.

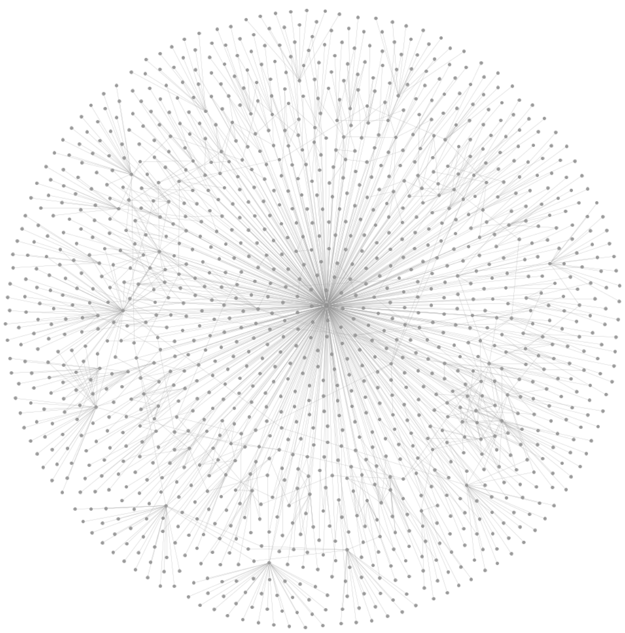

Fig. 55. Output of Fruchterman-Reingold Algorithm

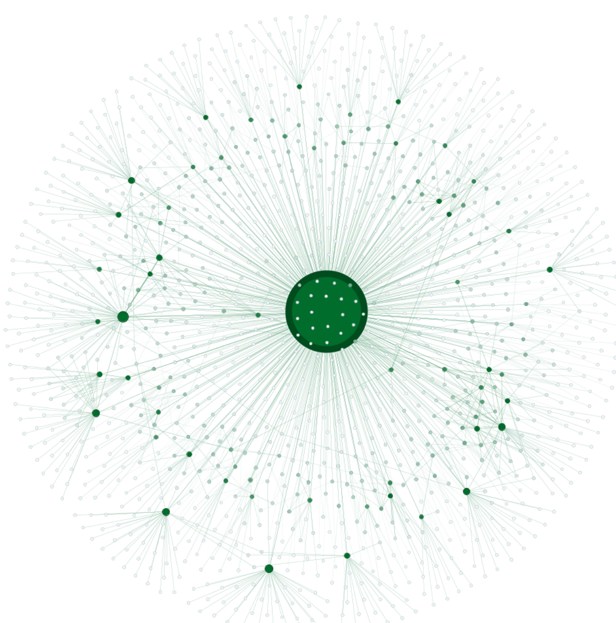

Fig. 56. Mapping the degree of nodes to a color and Size

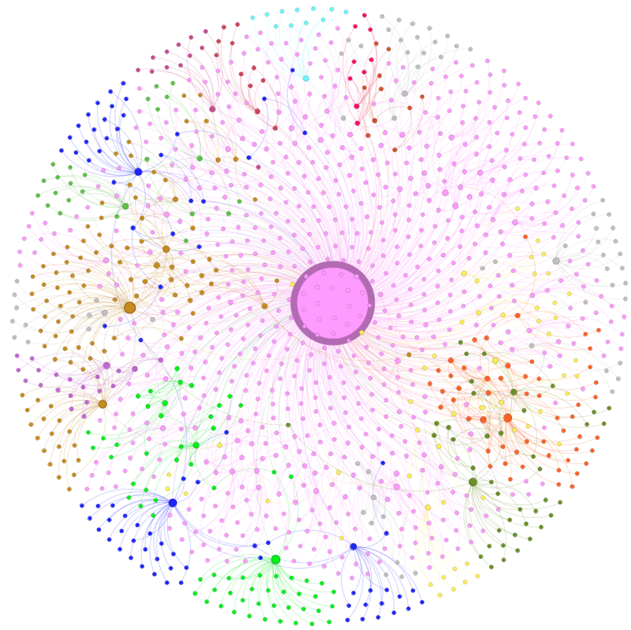

Fig. 57. Identifying clusters through modularity in the Fruchterman-Reingold algorithm

The output of this algorithm can be seen in Fig.55.

In Fig.56. we applied color and size rankings to nodes based on degree using Gephi's ranking options. Nodes with higher degrees are darker and larger.
In Fig.57, we used Gephi's modularity feature to detect communities within the network.

The Fruchterman-Reingold layout, spreads nodes out more evenly across the space as compared to the other layouts in which the edges are cluttered. The nodes with fewer connections are placed farther away from the center, indicating their lower influence within the network.

Yifan Hu

The Yifan Hu layout algorithm [3] belongs to the category of force-directed algorithms, which includes the Force Atlas and Fruchterman Reingold algorithms. This algorithm is faster than the Force Atlas algorithm because of the way it optimizes the overall internode repulsions in the network. The algorithm uses both attractive forces and repulsive forces to arrange the graph.
Gephi's Yifan Hu layout algorithm parameters are Optimal Distance, Relative Strength, Theta, Convergence Threshold, Adaptive Cooling, Step Ratio.

The output of this algorithm can be seen in Fig.58.

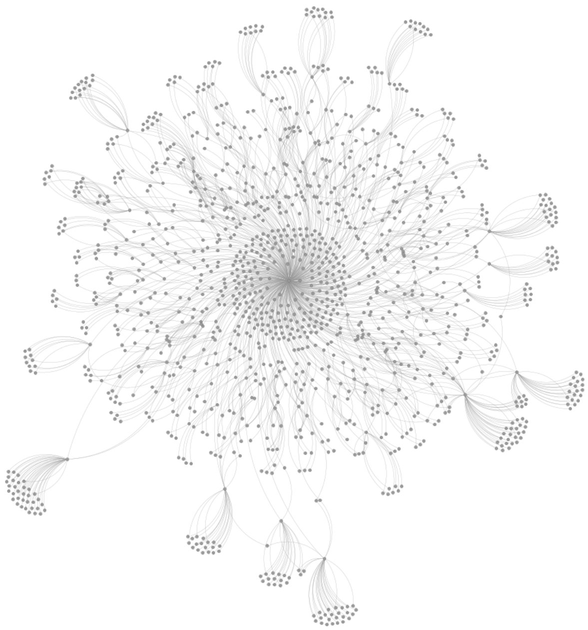

Fig. 58. Output of Yifan Hu layout algorithm

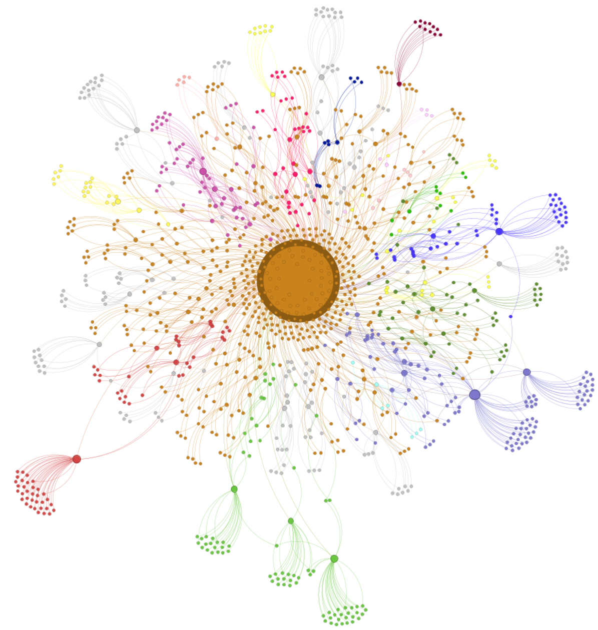

Fig. 60. Identifying clusters through modularity in the Yifan Hu algorithm

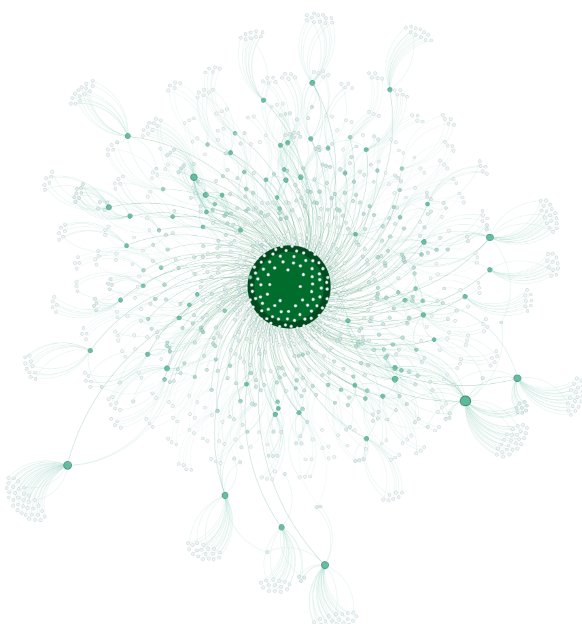

Fig. 59. Mapping the degree of each node to a color and Size in Yifan Hu layout

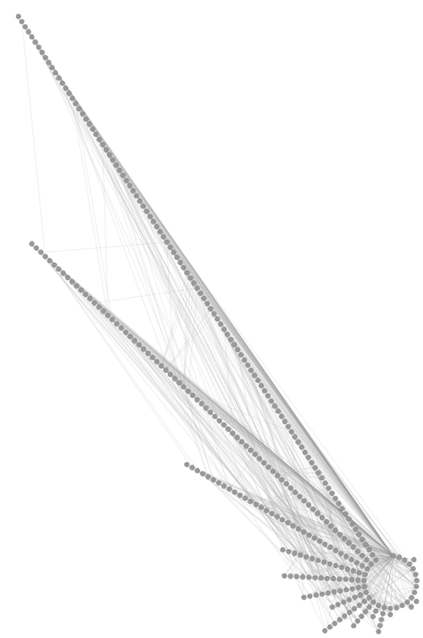

Fig. 61. Output of Radial axis layout Algorithm

In Fig.59, we utilized Gephi's ranking options to adjust the color and size of nodes based on their degree. Nodes with higher degrees are displayed with a darker color and a larger size, highlighting their importance within the network.

In Fig.60, we applied Gephi's modularity feature to identify communities within the network.

Radial Axis Layout

The Radial Axis Layout organizes nodes into groups and arranges these groups along axes, radiating outward from a central circle. Groups are generated using a metric or an attribute.

We have not included the nodes with degree less than 3 in the figures since it looked too cluttered. We have different parameters for this layout in Gephi like node sizes, grouping the nodes, arranging the nodes.
The output of this layout can be seen in Fig.61.

In Fig.62 , different communities can be observed using different colors , which is being done with the help of Gephi's modularity feature.

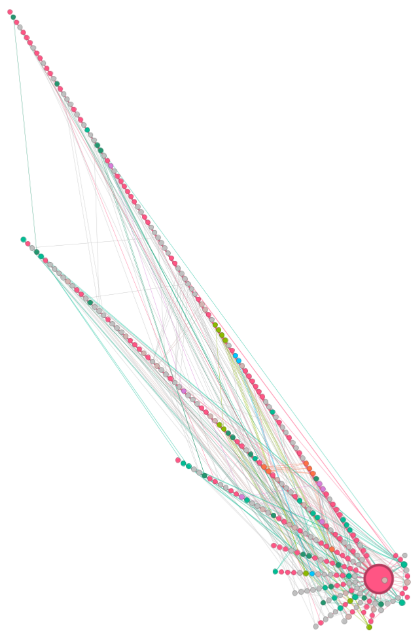

Fig. 62. Identifying clusters through modularity in Radial axis layout

With a large number of nodes, the Radial Axis Layout looks cluttered, making it hard to see individual nodes and connections. Limited space along each axis can cause nodes and edges to overlap, making the layout harder to read. This layout also unintentionally suggest a hierarchy, which misrepresents the relationships in given network.

OpenOrd

This algorithm processes an undirected, weighted graph to identify clusters efficiently. Based on Fruchterman-Reingold ,It follows a step-by-step schedule (liquid, expansion, cool-down, crunch, and simmer) that helps organize clusters, and it shortens long edges to keep clusters distinct. We have different parameters for this layout in Gephi like Random seed, iterations, fixed time, Edge Cut.

We have tried to map communities using different colors using Gephi's modularity feature.
The layout draws similar nodes closer together, but the clustering isn't as visually clear as in Force Atlas. Additionally, nodes within the same community often overlap, making them

Fig. 63. Output of OpenOrd Algoritm

hard to distinguish.
The output can be seen in Fig.63
In Fig.64 , different communities can be observed using different colors , which is being done with the help of Gephi's modularity feature

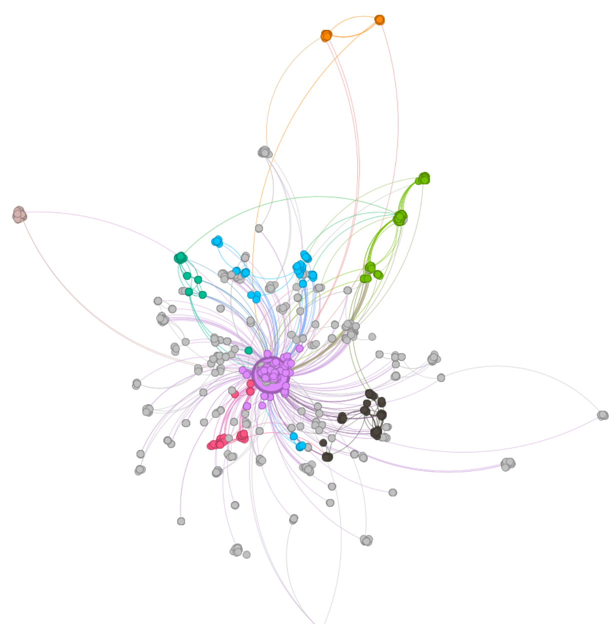

Fig. 64. Identifying clusters through modularity in OpenOrd Algoritm

Observation: We experimented with multiple layouts to visualize node relationships and community structures effectively. Force Atlas provided the clearest clustering, with popular nodes centrally positioned, making connections more readable. Other layouts, like Fruchterman-Reingold, grouped similar nodes but lacked visual clarity, and Radial Axis layout had limitations with overlapping edges in larger communities.

B. Treemap Plot

A treemap plot is a type of visualization that uses nested rectangles to represent hierarchical data. The size of each rectangle corresponds to a particular value, while the color of the rectangle can indicate another dimension or category of the data. In areas like finance or data analysis, treemaps are commonly used to display proportions of categories, such as sales by product or market share by company. This visualization allows for a compact, clear representation of complex hierarchical relationships, helping users easily identify patterns, outliers, and distributions within the data.

Fig 65. This is the page you will get once you run the HTML code, the website's name is TreeViz.

1) *Data Selection*: We selected a sample of data from the World Bank World Development Indicators dataset from Kaggle, focusing on countries globally. This dataset includes various statistics and attributes related to development indicators. By plotting treemaps, we aim to visualize these indicators for all the countries across the globe, providing insights into key development metrics. The analysis captures the distribution of various attributes, allowing us to observe trends and disparities across countries based on their performance in areas such as economy, health, education, and governance.

2) *Data Preprocessing*: The data was preprocessed by calculating the average of all attributes across all countries and grouping them by country name. The date column was removed, as it was deemed unnecessary for the analysis and would not contribute meaningful insights for the visualizations. This cleaned dataset allowed for more focused and effective visual representation of the data.

3) *Implementation*: The treemap visualization uses HTML, CSS, and JavaScript with Plotly [10] for dynamic, interactive visualizations. Data is loaded from a CSV file, and column names are initially displayed in a treemap. Users can click on any attribute to explore values grouped into ranges, with each level displayed in a hierarchical format for deeper data insights. Limited inherent hierarchies mean only specific attributes are suited for multi-level plotting, but Plotly's flexible design and responsiveness make it ideal for this task. Custom CSS enhances the interface, styling buttons and interactive elements, while tested color schemes ensure clarity and distinction in the treemaps. A color gradient represents value scales, enhancing data readability and enabling intuitive insights into attribute distributions and hierarchies within the dataset.

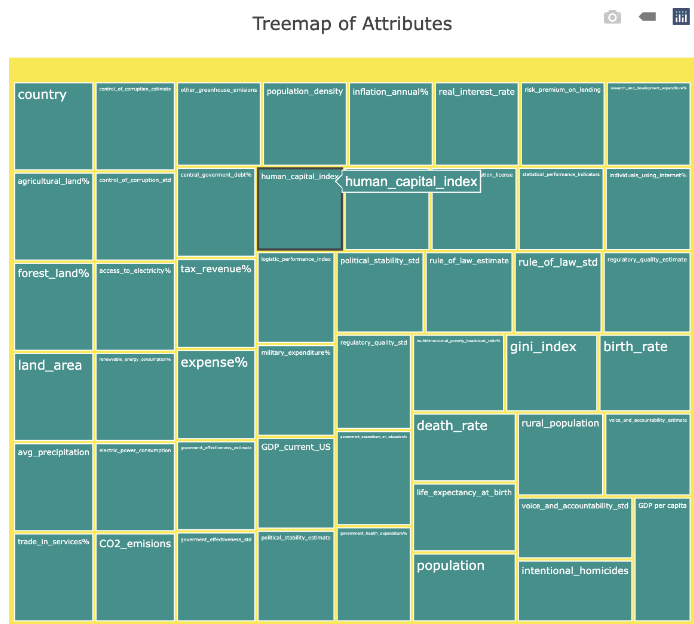

Fig 66. This image displays all the attributes from the selected CSV file as a treemap and illustrates how the hover functionality highlights information when the cursor is placed over a specific tile.

4) *User Interaction*: The treemap plot provides several interactive options for users beyond the HTML interface, as outlined below:
- Users can click on specific blocks within the treemap to drill down into ranges of values for a selected attribute, allowing them to explore a more detailed breakdown of the data hierarchy. For example, Fig.

shows the initial treemap of attributes, and Fig. displays data after selecting a particular attribute range.
- The color scale in the treemap enables visual differentiation of value ranges within selected attributes. By hovering over specific blocks, users can view detailed values and color-coded indicators that help distinguish high, medium, and low values across the treemap (Fig.).
- Users have the ability to download the treemap plot as an image. This feature provides flexibility for users who wish to save or share the visualizations for further analysis or presentation purposes. The option is easily accessible through a dedicated download button placed on the interface.
- The treemap includes an interactive hovering feature that displays detailed information when users hover over any block representing a country. Users can toggle this hover feature on or off based on their preference, allowing for a more streamlined view when detailed data is not needed. The toggle is accessible via a simple button in the user interface, ensuring ease of use without overwhelming the display.
- Additionally, when users hover over a specific country within the treemap, they can view the percentage share that country occupies within the entire treemap, based on its value relative to the total value of all displayed countries. This percentage indicator provides context on the proportion of each country's data, allowing users to quickly understand its relative significance in the dataset.

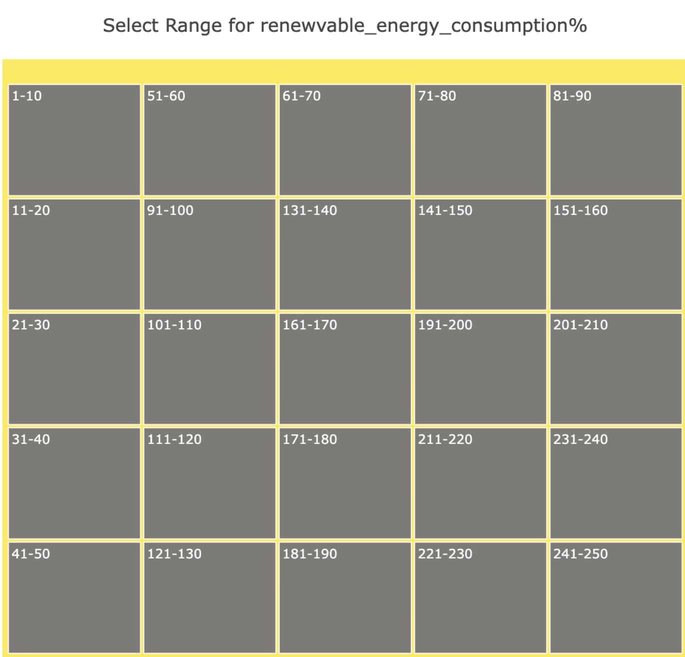

Fig 67. This image shows the treemap generated after a user selects a specific attribute and range for displaying countries. In this case, the attribute selected is renewable energy consumption

5) *Experiments*: Several experiments were carried out to optimize the treemap visualizations:
- **Hierarchical Data Exploration**: The treemaps were initially populated with the column names of the dataset. Clicking on a specific attribute within the treemap allowed users to explore data grouped by value ranges. This helped evaluate how effectively hierarchical structures could be visualized in terms of values, such as GDP, population, or other metrics.

- **Range-based Data Grouping**: Different value ranges (e.g., 1-10, 10-20) were tested to assess how well the treemap can handle data segmentation. The experiment involved sorting the dataset by the selected attribute and dividing the values into groups, allowing users to visualize data within specific intervals, particularly for top countries based on selected attributes.

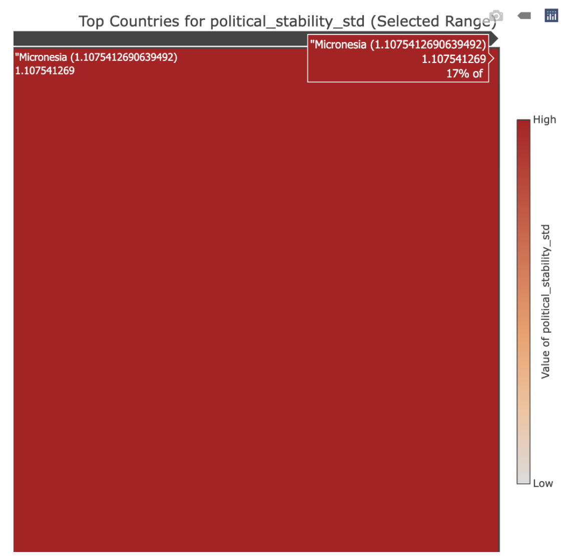

Fig 68. Clicking on a specific country in the treemap expands it to display as a standalone block, allowing the user to view it in full on the screen.

- **Color Scheme Optimization**: Various color scales (e.g., "Viridis" and "Cividis") were tested to assess their effectiveness in distinguishing different value ranges in the treemap. The experiments aimed to determine how well users could differentiate between high and low values using color intensity.
- **Data Filtering and Sorting**: The treemaps were tested with different sorting algorithms, such as sorting by the selected attribute in descending order to highlight the top 250 values. The filtering was done based on numeric values to ensure that only valid data points were displayed, improving the accuracy and responsiveness of the visualization.
- **User Interaction with Treemaps**: The interaction flow was tested by clicking on different blocks in the treemap, ensuring that the hierarchy and drill-down feature worked as expected. Users could click to explore more specific sub-categories or attributes within the

data, and the back button functionality was assessed to ensure smooth navigation between views.

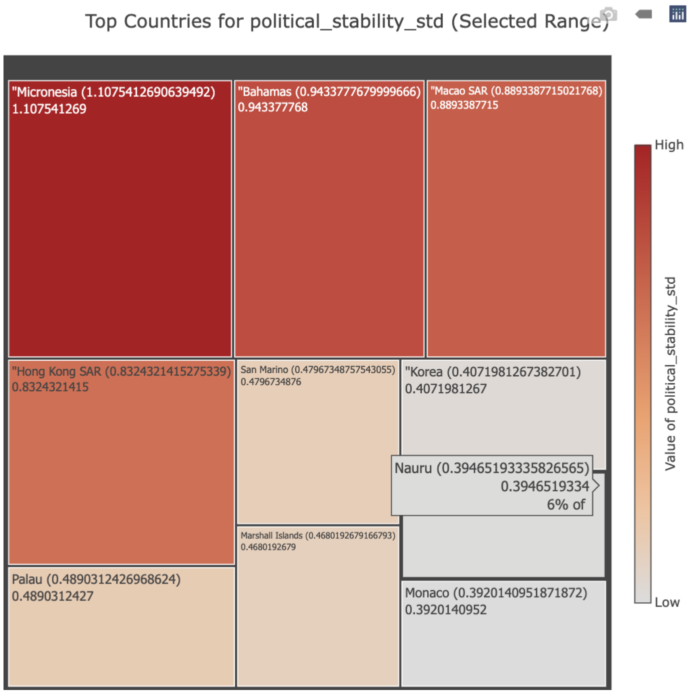

Fig 69. Hovering over a country reveals its name, value, and percentage share within the treemap, allowing users to make direct comparisons.

- Data Scale and Performance Testing: The performance of the treemap was evaluated with different dataset sizes, particularly focusing on the impact of large datasets (top 250 countries) on rendering speed and interactivity. This experiment helped ensure that the treemap could handle larger data volumes without significant delays or performance issues.

6) *Observation and Inferences*: The treemap visualization presented below offers a comprehensive overview of the distribution of various attributes across countries over the years. This interactive treemap allows users to explore the hierarchical structure of attributes, providing detailed insights into the economic, social, and political factors influencing each region. We have over 40 attributes but we are taking 4 different attributes for our visualisations. The treemap efficiently conveys the scale and proportions of various attributes such as GDP per capita, annual inflation, population, and military expenditure in different countries. Each block in the treemap represents a country, with the size of the block proportional to the respective attribute's value. Users can interactively navigate through the treemap to delve into specific countries and understand the nuances of the values within them.

- **Visualization 1 - GDP per Capita of Countries (1960-2022)**: The treemap visualization below offers a comprehensive overview of the distribution of GDP per capita across the top 10 countries in 2022. This interactive treemap allows users to explore the hierarchical structure of GDP per capita, with Monaco, Cayman Islands, and Liechtenstein topping the list. The treemap effectively conveys the stark contrast in economic development between these top-ranked countries and the rest of the world. For instance, Monaco's GDP per capita significantly exceeds the rest, reflecting its highly developed economy. By entering the Monaco node, users can explore how its GDP growth has remained strong over the years, contributing to its ranking. Similarly, the Cayman Islands and Liechtenstein also showcase impressive economic stability, reinforcing their high positions in global economic standings. **Analysis**: The treemap allows users to see that countries with smaller populations, such as Monaco, benefit from high GDP per capita due to their financial systems, tourism, and tax policies. The differences between the top three countries and others highlight the concentration of wealth in small, high-income nations.

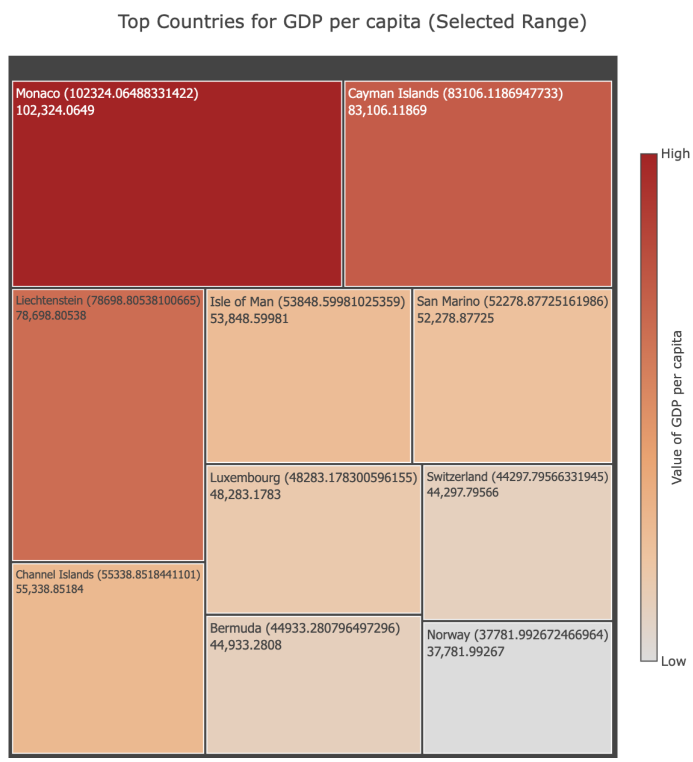

Fig 70. Visualisation 1 : Top 10 countries by average GDP per capita from 1960 to 2022

- **Visualization 2 - Annual Inflation of Countries (1960-2022)**: This treemap visualization delves into the distribution of annual inflation rates from 1960 to 2022 across countries ranked in the 51-60 range. Colombia, Jamaica, and Myanmar feature prominently. The treemap illustrates how inflation rates in these countries have fluctuated over the years, with notable spikes in some years. For example, Colombia and Jamaica exhibit periods of high inflation, which are likely tied to economic instability, while Myanmar's inflation also aligns with political and economic changes.

Analysis: By exploring the treemap nodes, users can observe the correlation between political events and inflation rates. In particular, high inflation rates in these countries during certain years suggest the impact of domestic instability and external economic factors. Users can also interactively explore how these countries' economies have struggled with inflation and identify the key periods when interventions or reforms were needed.

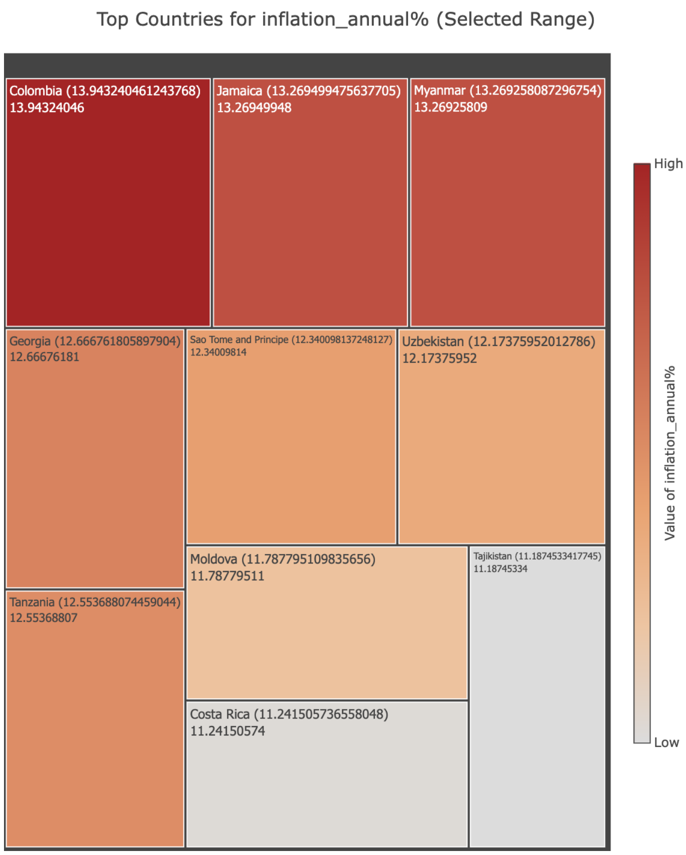

Fig 71. Visualisation 2 : Countries ranked 51-60 based on average annual inflation from 1960 to 2022.

- **Visualization 3 - Population of Countries (1960-2022)**: The treemap visualization for population growth, representing countries ranked 71-80, illustrates the significant demographic changes that have occurred over the decades. Argentina and Colombia are among the leading countries in this range. Argentina's population growth trajectory reflects its steady increase over the years, while Colombia shows similar trends, albeit with some fluctuations during certain periods.

Analysis: The size of the blocks represents the growing populations of these countries, and by interacting with the treemap, users can observe how Argentina and Colombia's populations have surged, reflecting both natural growth and migration patterns. This trend indicates the challenges these countries may face in terms of resource management, urbanization, and public health needs in the coming years. Additionally, sustained population growth in these regions can impact economic development, requiring advancements in infrastructure, housing, and employment opportunities to support expanding populations. The visualization provides a comparative snapshot, highlighting how the demographic landscape of these nations may shape future policy considerations around sustainability and social services.

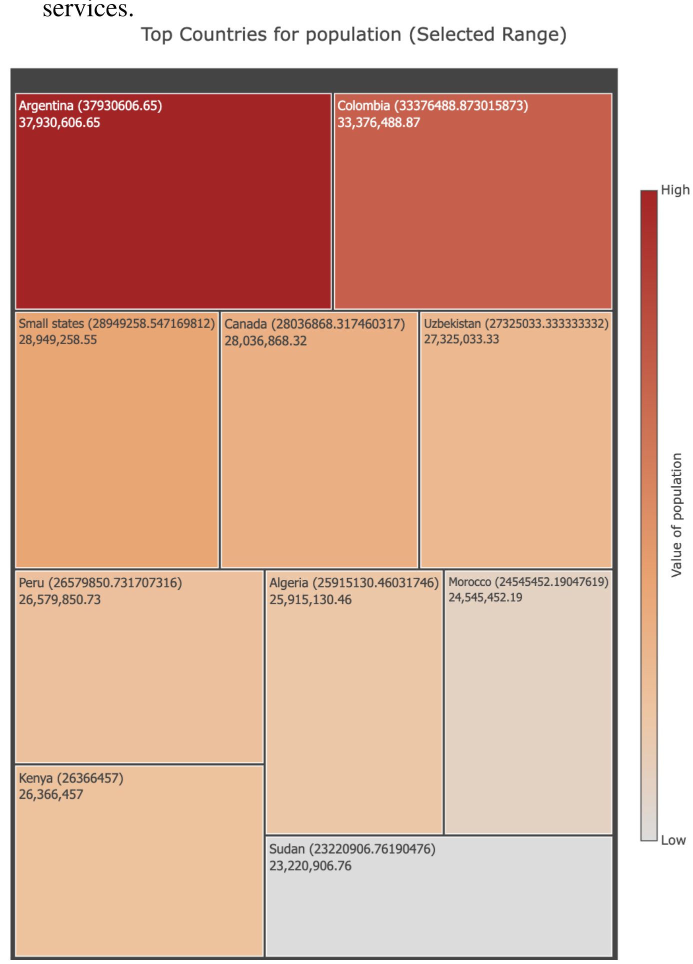

Fig 72. Visualisation 3 : Countries ranked 71-80 based on average population from 1960 to 2022.

- **Visualization 4 - Military Expenditure of Countries (1960-2022)**: The treemap visualization for military expenditure between 1960 and 2022 focuses on the countries ranked 21-30. Lebanon, Bahrain, and Singapore lead this group. The treemap highlights the fluctuations in military spending, with particular spikes in countries facing political instability or security concerns. For example, Lebanon's military expenditure saw an uptick during periods of conflict, while Bahrain's spending reflects its strategic investments in defense due to geopolitical tensions in the region. Similarly, Singapore's military spending is closely tied to its defense strategies in a region with significant security challenges.

Analysis: The treemap allows users to explore the varying military expenditures in these countries, showing how geopolitical concerns often influence defense budgets. The analysis reveals how countries in this range prioritize military spending due to external threats or internal political challenges, with specific focus on Lebanon, Bahrain, and Singapore's spending during key events.

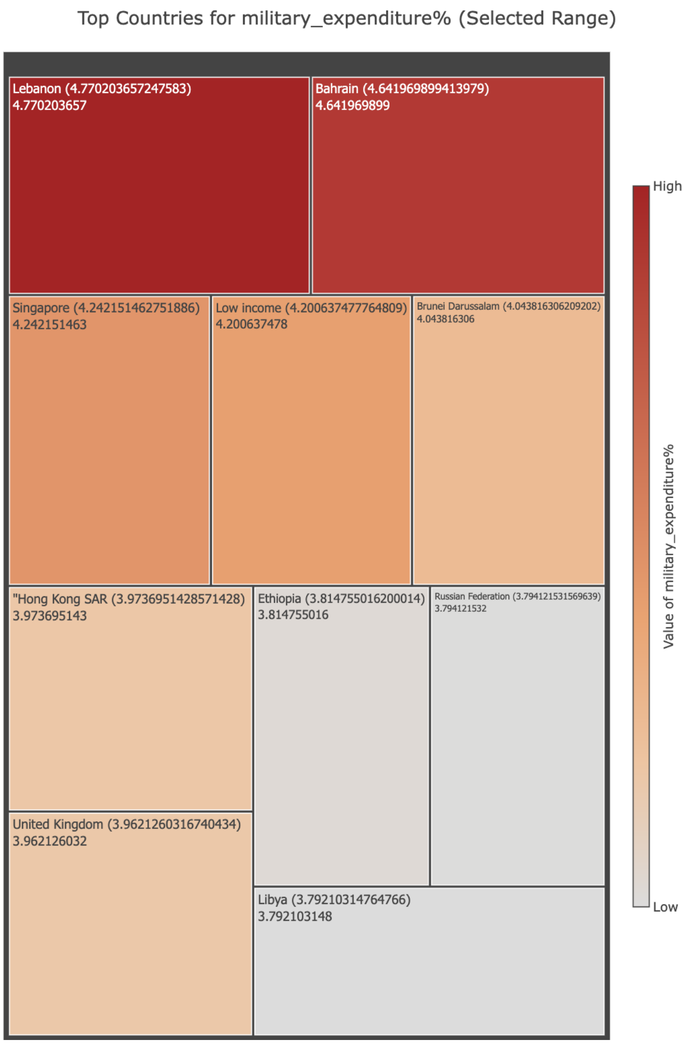

Fig 73. Visualisation 4 : Countries ranked 21-30 based on average military expenditure from 1960 to 2022.

C. Parallel Coordinates Plot

A parallel coordinates plot is a visualization technique used to display multivariate data. It consists of a set of parallel axes, each representing a different variable. Data points are plotted as lines that intersect each axis at the corresponding value. This visualization is particularly useful for exploring relationships between multiple variables and identifying patterns, trends, and clusters within the data.

Dataset: The same dataset as Assignment-1 i.e."World Bank World Development Indicators" sourced from kaggle was used for the Parallel Coordinates Plot (PCP). This dataset includes various statistics and attributes related to development indicators. We used a sample from this dataset covering a diverse range of countries. By utilizing parallel coordinates plots, we can simultaneously visualize multiple indicators for each country, facilitating cross-country comparisons across a variety of attributes. This approach enables us to observe patterns and contrasts in the data.

Preprocessing of Dataset:
For this analysis, we focused on environmental metrics such as emissions and energy consumption, selecting data from the year 2010. This choice provides a clear, concise view of environmental indicators without the need for additional preprocessing. All non-country regions and groups have been removed, leaving only selected countries. By examining data from a single year, we are able to make direct cross-country comparisons based on emissions and resource use, facilitating straightforward analysis and interpretation.

Implementation: The parallel coordinates plot (PCP) was created using HTML, CSS, and JavaScript, with Plotly.js for interactive plotting. Data is loaded from a CSV file, and users can select different attributes to visualize using checkboxes. Each country's data appears as a line that runs across parallel axes representing different attributes. Plotly's customization options make it easy to adjust the plot's appearance, such as changing axis scales and smoothing lines. The interface is styled with custom CSS to improve user experience, providing clear labels and an easy layout for selecting data and generating the plot.

The PCP includes axis reordering, allowing users to drag and rearrange axes to bring focus to desired comparisons between attributes. Additionally, brushing is incorporated to enable selection of specific ranges on any axis, making it easy to filter and emphasize relevant data points. Custom CSS improves the interface, and a color scale adds clarity by visually distinguishing attribute values. Plotly's interactivity further enhances data exploration, enabling detailed, dynamic analysis of environmental and infrastructure metrics across countries.

Experimentation: For parallel coordinates plot (PCP), several experimental steps were undertaken to optimize data representation and enhance user interactivity:

- **Attribute Selection and Ordering:** Different attribute combinations were tested to find sets that provide the clearest environmental insights. Users can select specific attributes to plot, adjusting the axes to create more meaningful comparisons between metrics. Additionally, reordering axes was tested to assess its impact on readability, enabling flexible exploration of attribute relationships.
- **Color Scale Testing:** Various color schemes (such as *Viridis* and *Plasma*) were experimented with to improve visual differentiation of attribute values across countries. The aim was to enhance clarity, allowing users to easily distinguish between low and high values for selected environmental and infrastructure attributes.
- **Brushing for Range-Based Filtering:** The PCP was tested with brushing capabilities, allowing users to focus on specific ranges within an attribute. This feature enables examination of subsets, such as countries with high or low emissions, while filtering out less relevant data points. Brushing experiments helped determine how this interaction could highlight specific patterns without overcrowding the visualization.
- **Axis Scaling and Normalization:** Different scaling techniques, such as normalizing attribute values, were explored to balance varied data ranges. This ensured that metrics with larger scales, like emissions or electricity usage, would not overshadow smaller-scaled metrics. Normalization experiments improved comparability across diverse indicators.

These experiments refined the PCP's interactive features, providing users with a flexible and intuitive tool for exploring and analyzing complex environmental and infrastructure trends across countries.

Here, we aim to analyze global environmental data, with a focus on understanding CO_2 emissions and certain related attributes.
Initially, all available countries were included to visualize CO_2 emissions and other relevant attributes. However, this resulted in a highly cluttered plot, as shown in Fig. 74, where the high number of overlapping lines obscured potential insights. This problem persists even when other colour maps are chosen.
To address this issue and improve clarity, we implemented the following preprocessing steps:

- Sorting by CO_2 Emissions: The dataset was sorted in descending order of CO_2 emissions. This ordering provided an immediate visual focus on countries with the most significant environmental impact.
- Filtering Countries: To reduce visual clutter, we initially selected the top 50 countries based on CO_2 emissions. Later, the dataset was further filtered to include only the top 10 countries, allowing for a more detailed analysis without overcrowding the plot.

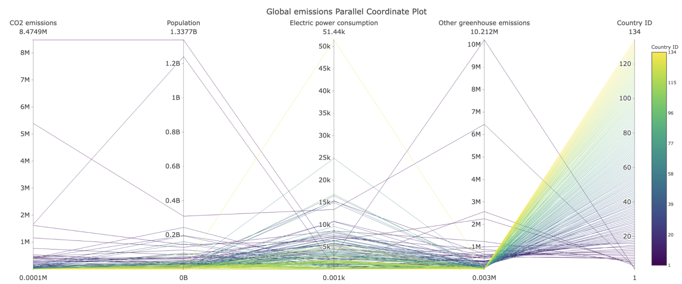

Fig. 74. The Plot is very cluttered when rows are not filtered

This gives us a better and uncluttered initial PCP layout. Fig.75 shows the initial layout of PCP after the filtering. The following attributes of the dataset are plotted:

- *Country :* The concerned country. Originally strings, using a 1-indexed label encoding for plotting.
- *CO_2 emissions :* The total amount of CO_2 emissions in (kt).
- *Other greenhouse emissions :* The total amount of other greenhouse gas emissions in (kt of CO_2 equivalent).
- *Electric power consumption :* The amount of electric power consumption in (kWh per capita).
- *Population :* The total population of a country.

The mapping from Country ID, as marked on the axis, to the corresponding country name is provided below the plot, as shown in Fig. 75.

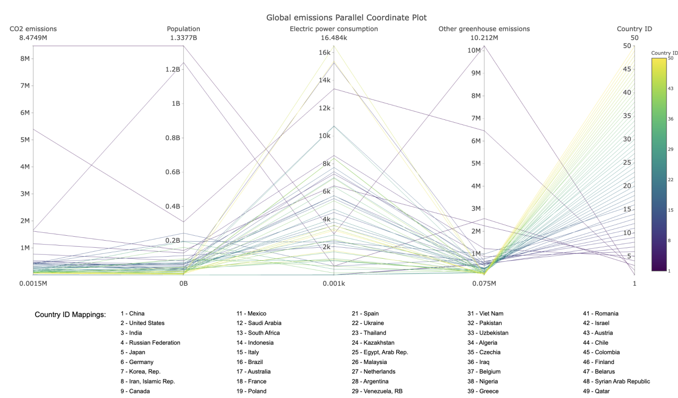

Fig. 75. Initial PCP layout after filtering for top 50 CO_2 emitting countries

Interactions : The plot is interactive to aid with the process of Visual Analytics. Axis brushing and Axis reordering can be used as is convenient for the user.
Axis reordering can also be used to observe direct correlations between two attributes. This is done via clicking

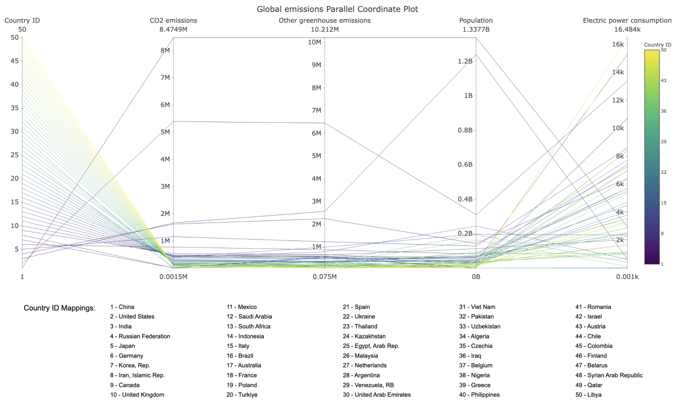

Fig. 76. Using axis reordering to study the correlation between two attributes

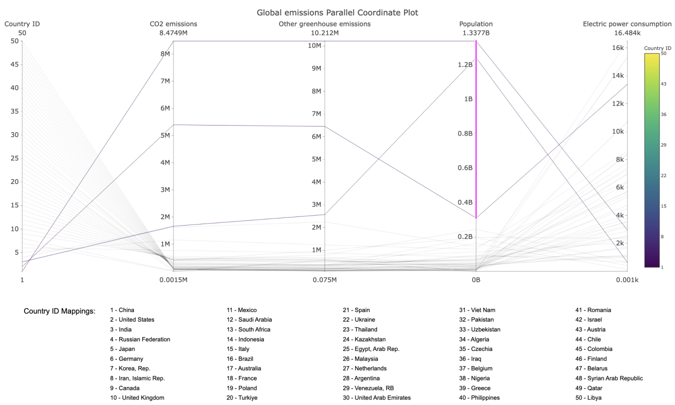

Fig. 77. Using axis brushing to highlight the top 3 most populous countries

and dragging the name of the attribute and moving it as desired. For example, in Fig.76 the relation between CO_2 emissions and other greenhouse gases emissions is more visually apparent after the axis reordering is done as compared to Fig.75

Axis Brushing can be used to filter out and select only certain ranges of attributes at once. This is done via clicking and dragging the mouse over the desired range(s) on the axis of any attribute. For example, Fig.77 shows how only certain Population values can be selected using Axis Brushing. Note that we can select multiple ranges within a single attribute or apply selections across multiple attributes simultaneously. This is demonstrated by examples in the following sections.

After initially filtering the data to display the top 50 countries by CO_2 emissions, we noticed that the plot still appeared somewhat cluttered. This was due to the significant variation in CO_2 emission values within this group—specifically, a few countries have extremely high emissions, while the remaining countries have considerably lower values. This wide range in emissions causes the high-emission countries to dominate the visual space, compressing the bristles for the other countries and making it difficult to see patterns clearly across the entire set.

To address this issue and enhance the plot's readability, we further refined the data to include only the top 10 countries. By focusing on these high-emission countries, we can more effectively observe relationships and trends without the visual congestion present in the previous broader dataset. Here, the Fig. 78 shows the PCP plot for concerned countries with axes already re-ordered to give best correlations.

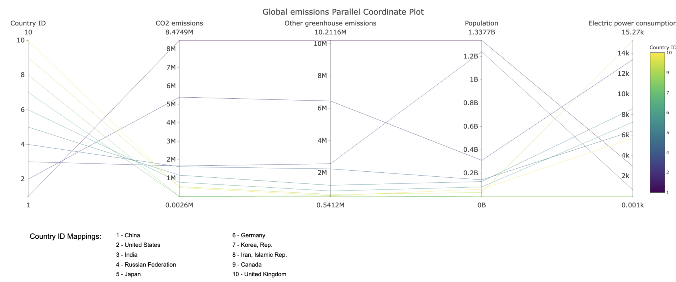

Fig. 78. PCP layout for top 10 CO_2 emitting countries

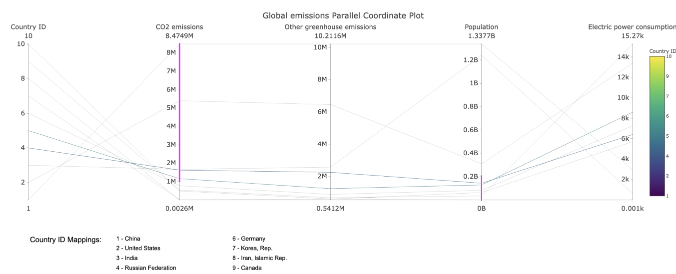

Fig. 79. Dual-axis brushing for precise selection across two attributes

Fig. 79 demonstrates the use of dual-axis selection, where two attributes—population and CO_2 emissions—are selected simultaneously. This allows users to define specific ranges for both attributes, helping to identify countries that fall within these selected ranges. By selecting ranges for both population and CO_2 emissions, users can explore how these two factors are related, and focus on the countries that meet both conditions. This method provides a clearer view of the correlations between population size and CO_2 emissions, making it easier to identify patterns or outliers in the data.

Furthur, Fig. 80 also demonstrates the use of axis brushing effectively. Here, the next 30 countries (after the top 10 in CO_2 emissions) are plotted, as they exhibit a more even distribution across CO_2 emission levels.

In this plot, as it can be seen, brushing is applied to highlight countries with CO_2 emissions between 100,000 and 300,000. This highlighting makes it easier to identify which countries fall within this specific emissions bracket and allows for a focused examination of related attributes for these selected countries.

Inferences: Using the interactions, the following inferences can be drawn using the PCP:

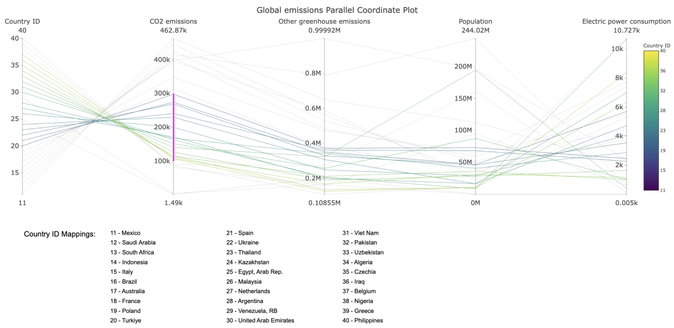

Fig. 80. Axis brushing to select a particular range of emission values

- Fig. 75 does not make it easy to see or interpret any clear correlations between the attributes.
- After reordering the axes, as seen in Fig. 76, certain patterns become more visible. A small group of high-emission countries stands out with much higher CO_2 values compared to others, making it clear that a few countries are contributing a large portion of emissions. The bristles connecting CO_2 emissions to other greenhouse gases emissions also reveal a positive correlation, meaning that countries with high CO_2 levels often have high levels of other greenhouse gases as well. Additionally, we start to see a vague indication that countries with larger populations are likely among those with the highest emissions.
- After performing axis brushing, as shown in Fig. 77, to highlight the top 3 most populous countries, we observe that these countries with the highest populations are also the top 3 contributors to global emissions. On checking the bristles and the ID mappings, we can identify these countries as China, USA and India.
- Fig. 78 clearly shows the general trend that countries with high CO_2 emissions are also major emitters of other greenhouse gases. Additionally, unlike the case in Fig. 26 with the top 50 countries, where emissions values below 1M were cluttered, only a few countries in the top 10 have emissions lower than 1M. It also reveals a clear relationship between emissions and population size, that countries with higher populations tend to have higher emissions.
- Using dual-axis brushing as shown in Fig. 79, we selected CO_2 emission values greater than 1 million and populations over 2 billion to identify countries that, while not among the top 3 most populous, still contribute significantly to global emissions. As expected, these countries turned out to be the 4th and 5th highest emitters. On checking the bristles and the ID mappings, we can identify these countries as Russia and Japan.
- Fig. 80 showcases how brushing can be extremely helpful in narrowing down specific data ranges when dealing with a large number of entities. In this case, by selecting CO_2 emission values between 100,000 and 300,000 and observing that no country with an ID lower than 20 is selected, we can conclude that none of the top 19 emitting countries contributed less than 300,000 metric tons of CO_2 in 2010. i.e. the 20th country marks the first instance where emissions fall below this threshold.

III. Contributions

The division of work is as follows:

- **Scientific Visualization**:
  - Quiver Plot: Ayush Arya Kashyap
  - Color Map: Uttam Hamsaraj
  - Contour Plot: Pranav Laddhad
- **Information Visualization**:
  - Node Link Diagram: Uttam Hamsaraj
  - Parallel Coordinate Plots: Pranav Laddhad
  - Tree Map: Ayush Arya Kashyap

Each team member independently completed their respective tasks, including data preprocessing, filtering and analysis, interpretation, description, and creating visualizations. We collaboratively discussed and selected the dates for SciViz, after which we proceeded with our individual tasks. This collective effort of every member has been compiled into this comprehensive report.

References

[1] https://support.noduslabs.com/hc/en-us/articles/4999513476370-Force-Atlas-Layout-Algorithm
[2] https://fiveable.me/key-terms/discrete-geometry/fruchterman-reingold-algorithm
[3] http://yifanhu.net/PUB/graph.pdf
[4] https://gephi.org/
[5] https://snap.stanford.edu/data/com-Amazon.html
[6] https://www.climatologylab.org/gridmet.html
[7] https://www.georgewbushlibrary.gov/research/topic-guides/hurricane-katrina
[8] https://matplotlib.org/
[9] https://pypi.org/project/Cartopy/
[10] https://plotly.com/javascript/