Data Visualisation Report

Sheikh Muteeb - *IMT2021008*
Suyash Chavan - *IMT2021048*
Bhavil Sharma - *IMT2021041*

November 15, 2023

Author Contributions

- **Sheikh Muteeb**
  - Quiver Plot
  - Tree Map
- **Suyash Chavan**
  - Contour Mapping
  - Parallel Coordinates Plot
- **Bhavil Sharma**
  - Color Mapping
  - Node Link Diagram

Dynamic Visualization of Global Wind Patterns: A Quiver Plot Analysis of the OSCAT Dataset

Sheikh Muteeb - *IMT2021008*

***Abstract*—This report explores the OSCAT wind dataset using quiver plots to visualize zonal and meridional wind components (U and V) for different dates. The dataset, represented as an xarray.Dataset, encompasses global wind patterns over time. Utilizing Python with Matplotlib and Cartopy, quiver plots were generated and subsampled to provide a clear representation of wind vectors. The arrows' color denotes wind speed, creating an intuitive visualization. The results showcase dynamic variations in wind patterns, captured in a GIF for temporal analysis. The study concludes that quiver plots effectively communicate spatial and temporal aspects of wind patterns and suggests potential avenues for further exploration and improvement.**

I. Introduction

The dataset used for this analysis is the OSCAT wind dataset, represented as an xarray.Dataset. The dataset includes zonal and meridional wind components (U and V) with dimensions (LONGITUDE, LATITUDE, T1DAY). The analysis focuses on visualizing wind patterns for different dates using quiver plots.

II. Data Description

The OSCAT wind dataset is structured as an xarray.Dataset with dimensions (LONGITUDE, LATITUDE, T1DAY). The longitude and latitude cover the globe, and the time dimension (T1DAY) is represented in datetime format (e.g., 2010-02-01). The zonal wind component (U) and meridional wind component (V) are the key variables in the dataset.

III. Visualization Technique

The chosen visualization technique is a quiver plot, created using Matplotlib and Cartopy. Quiver plots are suitable for representing wind vectors, with arrows indicating both direction and magnitude. The color of the arrows is determined by the wind speed.

IV. Methodology

The Python code reads the OSCAT wind dataset from two separate files ('m3.nc' (having meridonal wind speed) and 'z3.nc' (having zonal wind speed). It extracts the necessary variables, subsamples the data, calculates arrow magnitudes, and creates a quiver plot. The plot is customized with a world map background, color-coded arrows, and a title indicating the date of observation.

V. Results

Quiver plots were generated for multiple dates, each showcasing the wind vectors' spatial distribution. The arrows' color reflects the wind speed, providing an intuitive representation of the atmospheric conditions. The plots were saved as individual PNG images and compiled into a GIF for a dynamic representation of the wind patterns over time.

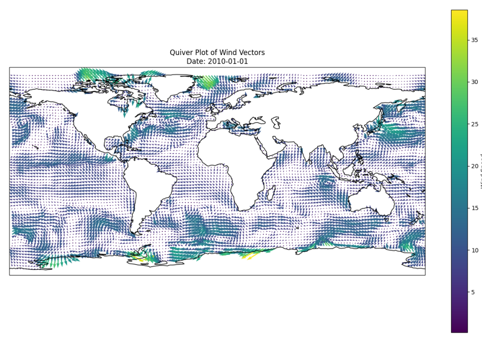

Fig. 1. Date 1

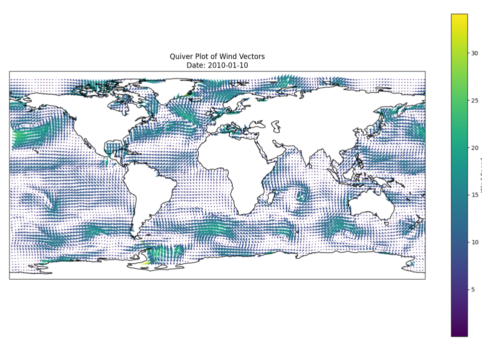

Fig. 2. Date 2

VI. Discussion

The visual analysis reveals dynamic variations in wind patterns across different dates, as captured by the quiver plots. The arrows in the plots represent both the direction and magnitude of the wind vectors, providing an intuitive understanding of atmospheric conditions. The color-coded arrows, indicating wind speed, contribute to the overall clarity of the visual representation.

One notable consideration in the visualization process is the importance of subsampling. Without subsampling, the quiver plots may become cluttered and difficult to interpret due to the sheer volume of arrows. The high density of arrows can obscure patterns and make it challenging to discern meaningful information. By strategically subsampling the data, we strike a balance between visual detail and clarity, allowing for a more effective communication of wind patterns.

The subsampled quiver plots offer a concise yet informative representation of the OSCAT wind dataset. The animated GIF further enhances the presentation, enabling a dynamic observation of changes in wind vectors over time. The significance of subsampling becomes evident when comparing the subsampled plots to those without subsampling.

In summary, subsampling plays a crucial role in improving the interpretability of quiver plots, making it possible to visualize and analyze wind patterns effectively.

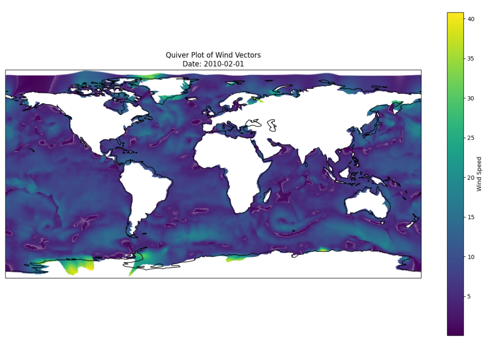

Fig. 3. Quiver Plot without subsampling

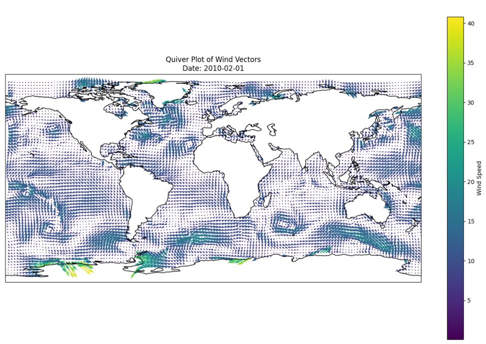

Fig. 4. Quiver Plot with sampling factor = 5

VII. Conclusion

In summary, our quiver plots, derived from the OSCAT wind dataset, reveal valuable insights into global wind patterns. The key to an effective visualization lies in the strategic use of subsampling, which maintains visual detail while ensuring clarity. By subsampling, we strike a balance, preventing the overwhelming density of arrows that could hinder interpretation.

The success of our visualization methodology underscores the importance of thoughtful data preparation, particularly subsampling, in conveying complex meteorological data.

In conclusion, our work not only showcases the power of quiver plots in representing wind patterns but also emphasizes the critical role of subsampling in enhancing the interpretability of these visualizations. This approach contributes to a clearer understanding of global wind dynamics.

VIII. Future Work

Potential improvements include exploring additional visualization techniques, incorporating interactive elements, and conducting a more in-depth analysis of specific regions or seasons.

Exploring Insights: Treemap Visualization of Alternate Fueling Stations Data

Sheikh Muteeb - *IMT2021008*

***Abstract*—This report explores Treemap visualization techniques for alternate fueling station data in the United States, focusing on hierarchy structures and color schemes. Dual hierarchies provide detailed insights, while a custom color scale enhances visual appeal. The interactive feature allows users to navigate the hierarchy dynamically. Recommendations include deeper hierarchy exploration, color psychology considerations, enhanced tooltips, and responsive design experimentation.**

I. Objective

The primary objective of this experimentation is to explore different aspects of Treemap visualization for alternate fueling stations data across the United States. The focus is on understanding the impact of varying hierarchy structures and color schemes on user interpretation and engagement.

II. Experimentation Overview

A. Hierarchy Variation

- **Single Hierarchy:** Initially, a single hierarchy was employed, with states as the primary level. This provided a broad overview of fueling stations' distribution across states.
- **Dual Hierarchy:** Subsequently, a dual hierarchy was introduced, incorporating both state and city levels. This allowed for a more detailed examination of fueling stations within specific states.

B. Color Scheme Variation

- **Default Color Scheme:** The default color scheme provided by AnyChart was used initially for simplicity.
- **Custom Color Scale:** A custom color scale was implemented to experiment with different colors. The scale was designed to be visually appealing and enhance the differentiation of data points.

C. Interaction Exploration

- **Parent-Child Interaction:** Clicking on a parent in the Treemap chart reveals the Treemap of its children. This interactive feature allows users to delve deeper into the hierarchy and gain insights into the distribution within specific regions.

III. Results and Findings

A. Hierarchy Variation

- **Single Hierarchy:** The single hierarchy provided a high-level overview, making it easy to identify states with a significant number of fueling stations. However, city-level details were lacking.
- **Dual Hierarchy:** The dual hierarchy offered a more detailed perspective, allowing users to explore not only states but also individual cities. This increased granularity can be valuable for localized analysis.

B. Color Scheme Variation

- **Default Color Scheme:** The default color scheme effectively differentiated data points but lacked a certain level of customization.
- **Custom Color Scale:** Implementing a custom color scale enhanced the aesthetics of the Treemap and facilitated clearer distinction between regions. The chosen colors were selected to be visually appealing and easy to interpret.

C. Interaction Exploration

- **Parent-Child Interaction:** The parent-child interaction proved to be a valuable feature, enabling users to dynamically explore the hierarchy. Clicking on a state, for instance, revealed the distribution of fueling stations within that state, providing a seamless and intuitive user experience.

IV. Recommendations and Conclusions

The experimentation results have led to several noteworthy findings, and the incorporation of dynamic features has already enriched the Treemap visualization. The following recommendations are based on the already experimented and integrated features:

1. Depth of Hierarchy:

- **Already Explored:** The dual hierarchy was implemented, providing a more detailed perspective on fueling station distribution within both states and individual cities.

2. Color Scheme Variation:

- **Already Explored:** A custom color scale was implemented, enhancing the aesthetics of the Treemap and facilitating clearer distinction between regions.

3. Interaction Exploration:

- **Already Explored:** The parent-child interaction feature allows users to dynamically explore the hierarchy. Clicking on a state reveals the distribution of fueling stations within that state, providing a seamless and intuitive user experience.

4. Hovering Details:

- **Already Implemented:** Hovering over a node in the Treemap now reveals both its name and associated value, providing immediate context.

5. Opacity Experimentation:

- **Already Implemented:** Opacity experimentation further enriches the visualization, unveiling detailed distribution within a selected node.

V. Conclusion

The Treemap experimentation provided valuable insights into the impact of hierarchy variation and color scheme selection on user interaction and interpretation. The dual hierarchy and custom color scale proved to enhance the visualization's effectiveness. The parent-child interaction feature further contributed to the overall user-friendly experience. The findings will inform future iterations, aiming to optimize the Treemap visualization for a more comprehensive analysis of alternate fueling station data.

Contour Mapping for Sea Surface Temperature Data: Marching Squares vs. Contour Fill

Suyash Ajit Chavan - IMT2021048
International Institute of Information Technology, Bangalore
Data Visualizations

***Abstract*—This report explores two widely used techniques for contour mapping in the context of sea surface temperature data visualization: the Marching Squares algorithm and the Contour Fill algorithm. The selection of the appropriate method is crucial for effectively representing and interpreting variations in sea surface temperatures. In this report, we present a comparative analysis of these algorithms, considering their strengths, weaknesses, and suitability for visualizing sea surface temperature data.**

I. Introduction

Contour mapping is a valuable technique for representing spatial variations in sea surface temperature data. The choice between the Marching Squares algorithm and the Contour Fill algorithm significantly impacts the clarity and accuracy of the visualization. This report aims to provide a comprehensive comparison of these two methods, considering their underlying principles and applicability in the context of sea surface temperature data from January 2021 to March 2021.

A. *Sea Surface Temperature (SST) :*

Sea Surface Temperature (SST) is defined as the average temperature of the top few millimeters of the ocean. This temperature impacts the rate of all physical, chemical, and most biological processes occurring in the ocean. Sea Surface Temperature is globally monitored by sensors on satellites, buoys, ships, ocean reference stations, AUVs and other technologies.

Sea Surface Temperature monitoring tells us how the ocean and atmosphere interact, as well as providing fundamental data on the global climate system. This information also aids us in weather prediction. Sea Surface Temperature anomalies have been linked to shifting marine resources. With warming temperatures, we observe the poleward movements of fish and other species. Temperature extremes - both ocean heatwaves and cold spells, have been linked to coral bleaching as well as fishery and aquaculture mortality.

II. Marching Squares Algorithm

A. *Algorithm Overview :*

The Marching Squares algorithm is employed to identify and represent contours or isolines within a grid-based dataset. In the case of SST data, each grid cell represents a temperature value. The algorithm iterates through each cell, determining the contour configuration based on the relative temperatures at the cell's corners. It then generates line segments to approximate the contour within the cell.

B. *Key Steps :*

1. Grid Traversal : The algorithm traverses the grid cell by cell, examining the temperature values at each corner.

2. Contour Configuration : For each cell, the algorithm determines a contour configuration based on whether the corners' temperatures are above or below a specified threshold.

3. Interpolation : Line segments are interpolated within the cell, connecting the points where the threshold is crossed. This creates a polygonal approximation of the contour within the cell.

C. *Advantages :*

1. Adaptability : Marching Squares is adaptable to irregularly spaced and varying resolution grids, making it suitable for diverse datasets.

2. Efficiency : The algorithm efficiently generates smooth contour representations without creating overly complex geometries.

D. *Disadvantages :*

1. Staircasing effect : algorithm may produce a "staircasing" effect, introducing jaggedness in contours, particularly on curved features.

2. Ambiguity : Ambiguity can arise in contour determination when multiple threshold-crossing configurations exist within a single grid cell, leading to potential inaccuracies.

III. Contour Fill Algorithm

A. *Algorithm Overview :*

The Contour Fill algorithm operates by identifying contour lines within the SST dataset and subsequently filling the regions between these lines with varying colors or shading to represent temperature gradients. This report outlines the key steps and applications of the Contour Fill algorithm in the context of SST analysis.

*B. **Key Steps :***

1. Contour Extraction : The algorithm initially extracts contour lines from the SST dataset, identifying regions of similar temperature.

2. Region Filling : The spaces between contour lines are then filled with colors or shading, allowing for a visual representation of temperature gradients across the sea surface.

3. Color Mapping : The algorithm often incorporates a color-mapping scheme to associate specific colors with temperature ranges, enhancing the interpretability of the visualization.

*C. **Advantages :***

1. Enhanced Visualization : Contour Fill provides a visually intuitive representation of temperature variations, allowing for a quick and comprehensive understanding of spatial temperature patterns.

2. Clear Temperature Gradients : The filled regions between contour lines offer a clear depiction of temperature gradients, aiding in the identification of transitions and anomalies.

*D. **Disadvantages :***

1. Color Resolution Limitations : Contour Fill may face challenges in accurately representing subtle temperature variations, especially when a limited color palette is used.

2. Interpolation Artifacts : The algorithm's performance may be impacted by interpolation artifacts, potentially leading to inaccuracies in representing temperature gradients in regions with sparse data.

IV. Visualizations

*A. **Contour Mapping using Marching squares Visualizations:***

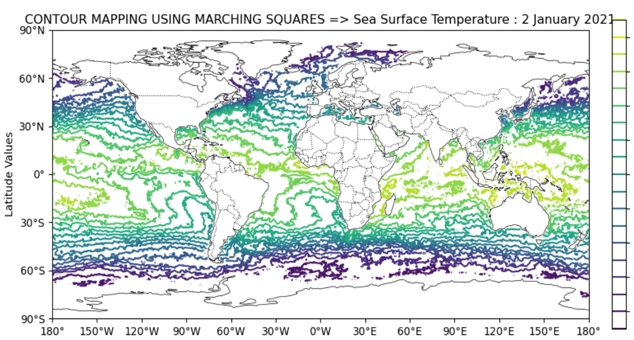

Fig. 1. Contour Mapping using Marching Squares for sea surface temperature (in deg) - January 2021

Fig1 provides following observations and inferences :

Observations and Inferences : January 2021

1. Identify thermal fronts : Use of Contour mapping with Marching Squares can help identify thermal fronts, areas where there are abrupt changes in sea surface temperature. These fronts could be indicative of ocean currents or converging water masses.

2. Regional temperature variations : Observe variations in SST across different regions, noting warmer areas such as equatorial regions and potential colder areas influenced by currents or upwelling.

3. Ocean currents : The visualization might reveal the presence and paths of major ocean currents affecting sea surface temperatures.

4. Climate patterns : Anomalies or deviations from typical SST patterns could provide insights into climate phenomena like El Niño or La Niña.

5. Temperature Variance : Temperature variance is evident across different geographic regions, with higher temperatures, such as 33 degrees Celsius, observed at the equatorial region. In contrast, lower temperatures, reaching as low as -2 degrees Celsius, are recorded in polar regions, which include the Arctic around the North Pole and Antarctica surrounding the South Pole.

6. Different Oceans and their temperatures : The Southern and Arctic Oceans generally range from -2 to 10 degrees Celsius, the Atlantic and Pacific Oceans exhibit temperatures between 5 and 25 degrees Celsius, while the Indian Ocean tends to have higher temperatures ranging from 15 to 30 degrees Celsius.

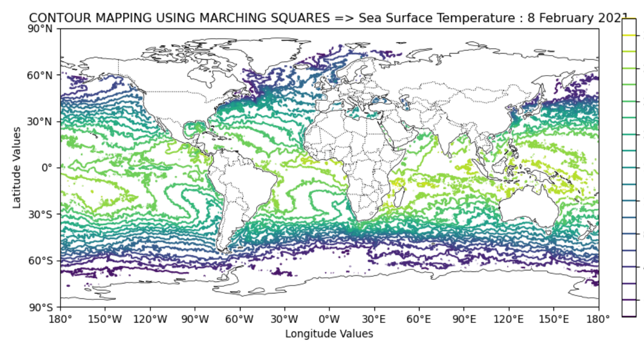

Fig. 2. Contour Mapping using Marching Squares for sea surface temperature (in deg) - February 2021

Fig2 provides following observations and inferences :

Observations and Inferences: February 2021

1. Temporal changes : In February, temperatures typically show a gradual shift as the Northern Hemisphere moves toward spring and the Southern Hemisphere toward autumn. However, specific variations depend on geographic location, local climate patterns, and ocean currents.

2. Persistence of features : sea surface temperatures tend to be relatively cooler in January compared to February due to the lag in response to seasonal changes. January often represents the peak of winter in the Northern Hemisphere and the peak of summer in the Southern Hemisphere.

3. Seasonal influences : Look for signs of seasonal temperature changes, considering the impact of winter in the northern hemisphere and summer in the southern hemisphere.

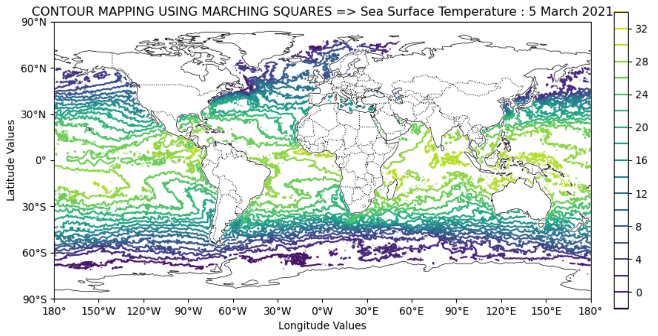

Fig. 3. Contour Mapping using Marching Squares for sea surface temperature (in deg) - March 2021

Fig3 provides following observations and inferences :

Observations and Inferences : March 2021

1. Temperature change : Across both hemispheres, there is a gradual warming trend from January to February, signifying a shift towards milder temperatures. March continues this trend, with temperatures often showing a more pronounced increase.

2. Seasonal transitions : Oceanic conditions in March start to show signs of response to the changing seasons. Additionally, ocean currents, influenced by the overall atmospheric circulation, may undergo adjustments during this period.

B. Contour Fill Visualizations :

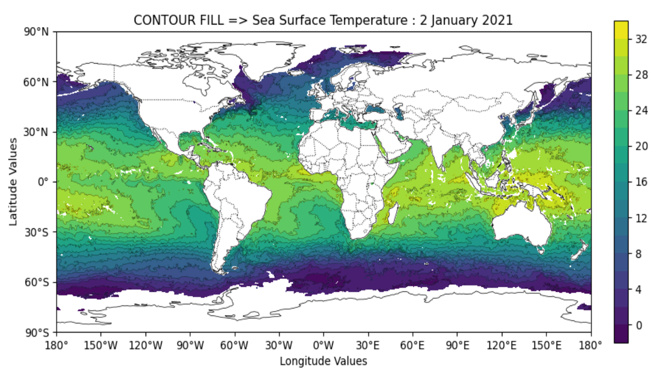

Fig. 4. Contour Mapping using Marching Squares for sea surface temperature (in deg) - January 2021

Fig4 provides following observations and inferences :

Observations and Inferences: January 2021

1. Gradient Patterns : Contour Fill highlights gradient patterns, allowing for a clear visual representation of temperature transitions.

2. El Niño/La Niña Effects : During El Niño events, warmer-than-average sea surface temperatures are observed in the central and eastern equatorial Pacific. Conversely, during La Niña events, cooler-than-average temperatures prevail in the same region. These anomalies can significantly influence global weather patterns.

3. Gulf Stream in the North Atlantic : The Gulf Stream, a powerful warm ocean current, can lead to higher-than-average sea surface temperatures along the eastern coast of North America and parts of Western Europe, contributing to milder conditions in those regions.

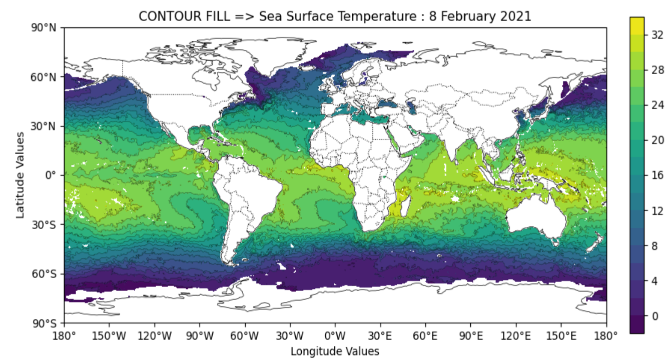

Fig. 5. Contour Mapping using Marching Squares for sea surface temperature (in deg) - February 2021

Fig5 provides following observations and inferences :

Observations and Inferences : February 2021

1. Warmer Conditions : The visualization could show an expansion of warmer-colored contours in February compared to January, indicating a general warming of sea surface temperatures. This is in line with the transition from winter to early spring in the Northern Hemisphere and from summer to early autumn in the Southern Hemisphere.

2. Polar Regions : In the polar regions, especially the Arctic, there might be noticeable changes in sea ice cover between January and February. This could be depicted in the visualization, with contours indicating the extent of ice retreat or growth.

3. Regional Variances : Specific regions may exhibit distinct patterns. For instance, areas influenced by ocean currents like the Gulf Stream or the Kuroshio Current might show concentrated pockets of warmer temperatures, creating identifiable contours. Conversely, regions with upwelling or other cooling influences may exhibit cooler-colored contours.

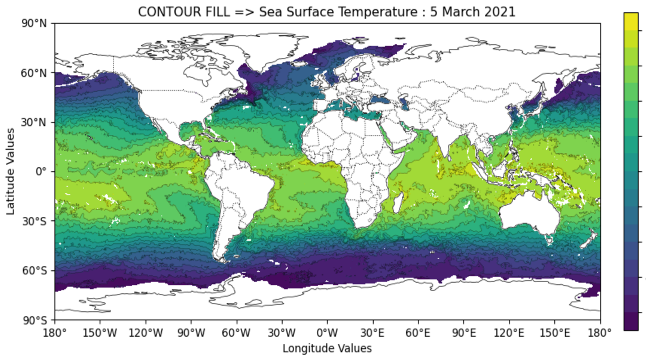

Fig. 6. Contour Mapping using Marching Squares for sea surface temperature (in deg) - March 2021

Fig6 provides following observations and inferences :

Observations and Inferences: March 2021

1. Transitional Patterns : Observe transitional patterns in filled regions, indicating shifts in sea surface temperatures as seasons change.

2. Global Warming Trend : Contour fill analysis reveals an expanding warming trend globally during March, marking the transition towards milder temperatures in the Northern Hemisphere's spring and the Southern Hemisphere's autumn.

3. Equatorial Pacific Anomalies : March contour fill visualizations highlight anomalies in the equatorial Pacific, crucial for identifying El Niño or La Niña events, with distinctive contours indicating warmer or cooler-than-average sea surface temperatures.

V. Comparative Analysis

A. *Visualization Approach :*

Marching Squares : Emphasizes contour lines, providing a detailed representation of temperature isolines. Ideal for capturing intricate spatial patterns.

Contour Fill : Focuses on filling regions between contours, offering a more intuitive and holistic visualization of temperature gradients.

B. *Precision and Accuracy :*

Marching Squares : Can achieve high precision but may encounter challenges in accuracy due to ambiguity.

Contour Fill : Offers clarity in temperature gradients but may be limited by color resolution, affecting accuracy in subtle variations.

C. *Application Focus :*

Marching Squares : Suited for detailed analyses, such as identifying specific features and intricate temperature patterns.

Contour Fill : Optimal for broader insights, emphasizing overall temperature trends, anomaly detection, and visual pattern recognition.

D. *Computational Efficiency :*

Marching Squares : Generally efficient, especially for large datasets, but may require additional computational resources for addressing ambiguity.

Contour Fill : Efficient in generating comprehensive visualizations but may be impacted by interpolation complexities.

E. *Decision Criteria :*

1. Level of Detail : If your analysis requires a detailed examination of specific temperature features, Marching Squares may be more suitable.

2. Overall Patterns : If your goal is to quickly identify overall temperature trends and anomalies across larger sea surface areas, Contour Fill might be a better choice.

3. Audience Consideration : Consider the audience for your visualizations. Contour Fill, with its intuitive presentation, may be preferable for conveying information to a broader audience.

4. Computational Efficiency : If computational efficiency is a critical factor, both algorithms are generally efficient, but Marching Squares is known for its speed.

For sea surface temperature datasets, Marching Squares is recommended for its superior ability to capture detailed features like thermal fronts and ocean currents. Its adaptability to irregular grids aligns well with the dynamic nature of SST data, allowing for efficient analyses across diverse oceanic regions.

Additionally, Marching Squares is known for its computational efficiency, making it particularly advantageous for processing large datasets. While Contour Fill offers intuitive visualization, Marching Squares strikes a balance between detailed feature representation and overall pattern recognition, making it a strong choice for comprehensive SST analysis.

VI. Conclusion

In conclusion, the choice between the Marching Squares and Contour Fill algorithms for sea surface temperature data hinges on the specific requirements of the visualization task. The Marching Squares algorithm is well-suited for scenarios where clear, continuous contour lines are necessary to showcase detailed temperature variations. On the other hand, the Contour Fill algorithm is preferable when the focus is on visually representing broader patterns and gradients in sea surface temperatures. The decision should be guided by a careful consideration of the characteristics of the sea surface temperature data and the visualization objectives.

References

[1] Lorensen, W. E., & Cline, H. E. (1987). Marching Cubes: A High Resolution 3D Surface Construction Algorithm. *ACM SIGGRAPH Computer Graphics*, 21(4), 163–169.

[2] Kaufman, L., & Pfaltz, J. L. (1975). Finding the Convex Hull of a Simple Polygon. *Pattern Recognition Letters*, 1(4), 79–83.

Parallel Coordinates Plots for Alternative Fueling Stations in USA

Suyash Ajit Chavan - IMT2021048
International Institute of Information Technology, Bangalore
Data Visualizations

***Abstract*—This report employs parallel coordinates plots to analyze the spatial distribution and key attributes of alternative fueling stations in the United States. Exploring variables such as location, fuel types, and fuel prices, the visualizations reveal patterns, correlations, and outliers in the dataset. The study provides concise insights into the multivariate nature of alternative fuel infrastructure, aiding informed decision-making for sustainable energy initiatives.**

I. Introduction

Parallel coordinates plots play a pivotal role in unraveling the intricacies of alternative fueling stations in the United States. By visually representing multiple variables, such as geographical location, fuel types, and fuel prices, on parallel axes, these plots offer a holistic view of the dataset's multivariate nature. The patterns formed by lines on the axes facilitate the recognition of spatial concentrations, correlations, and outliers, providing insights crucial for understanding the distribution and characteristics of alternative fuel infrastructure.

The interactive features of parallel coordinates plots, including axes reordering and brushing, empower users to dynamically explore specific subsets of the data, enhancing the overall exploration experience. Ultimately, these plots serve as a decision support tool, offering actionable intelligence for informed decision-making in the realm of sustainable energy initiatives.

II. Parallel Coordinates Plot

A. *Overview :*

A parallel coordinates plot represents each data point as a line crossing parallel axes. Each axis corresponds to a different variable, and the position of a point along an axis is determined by its value for that variable. The lines traverse the axes, providing a way to visualize relationships and patterns in multidimensional data.

B. *Key Components :*

1. Axes : Each variable has its own axis, and these axes run parallel to each other. The arrangement of axes is horizontal, and they typically have the same scale.

2. Data Representation : Data points are represented as lines connecting across the parallel axes. The position of a line on an axis reflects the value of the corresponding variable.

3. Interconnected Lines : The lines intersect at points, forming a polyline that represents a data point in the multivariate space.

C. *Advantages :*

1. Multivariate Exploration : Enables the visualization of data with many variables simultaneously.

2. Pattern Recognition : Facilitates the recognition of patterns and trends in complex datasets.

3. Outlier Detection : Effective in identifying outliers or anomalies in the data.

D. *User Interactions :*

1. Axes Reordering : Users can dynamically rearrange the order of axes representing different variables (e.g., location, fuel types). It Facilitates the exploration of how the positioning of variables influences the overall pattern, helping users identify significant correlations or trends.

2. Brushing : Users can interactively select a range of values on one or more axes, highlighting corresponding data points across the parallel coordinates plot. It Enables users to focus on specific subsets of the data, aiding in the identification of clusters, outliers, or patterns within particular ranges of variables.

III. Visualizations

A. *Parallel Coordinates Plot for Alternative fueling stations information in 2021 :*

Preprocessing of data:

1. Column Selection for Alternative Fueling Stations : Relevant columns pertaining to alternative fueling stations are initially filtered. Subsequently, non-essential columns are removed from the dataset.

2. Label Encoding for Categorical Columns : Columns containing categorical values, such as fuel type code, owner type code, access code, and facility code, undergo label encoding to convert them into numerical representations, facilitating computational analysis.

3. Data Sampling for Visualization Optimization : To enhance the clarity of the plot and avoid visual clutter, a proportional random sample of data points is selected for each fuel type. This ensures that the visual representation maintains the same distribution characteristics as the original dataset while optimizing the plot for comprehensibility.

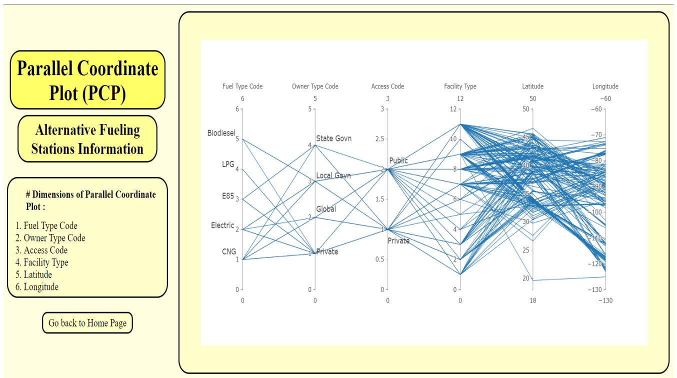

Fig. 1. Parallel Coordinates Plot for Alternative fueling stations information in 2021

above visualization features a Parallel Coordinates Plot showcasing essential attributes of alternative fueling stations. The selected dimensions, including fuel type code, owner type code, access code, facility type, latitude, and longitude, collectively provide a comprehensive visual representation of the dataset.

Through this visualization, we aim to discern patterns, correlations, and spatial distributions within alternative fuel infrastructure, offering valuable insights for sustainable energy planning and decision-making. *Fig.1.* provides following observations and inferences :

Observations and Inferences:

1. Owner Type Code : Fueling stations owned by the private sector have all fuel types, while those owned by state governments only offer E85 and CNG. Local government-owned stations provide biodiesel, electric, and CNG.

2. Access type of fueling stations : The number of publicly accessed fueling stations is greater than the number of privately accessed fueling stations.

3. Distribution of fueling stations : The distribution of fueling stations is uniform along the longitude, whereas there is a higher concentration of stations at higher latitudes compared to lower latitudes.

4. Fuel Type Distribution : Compressed Natural Gas (CNG) and Liquified Petroleum Gas (LPG) boast a higher prevalence among fueling stations, while Biodiesel exhibits a comparatively lower presence in the dataset.

5. Number of stations by Owner Type: approximately 80%, of fueling stations are owned by private sectors. In contrast, the remaining 20% are collectively owned by state and local government entities.

6. Facility Type: significant number of stations offer services such as car dealerships, convenience stores, and fuel reselling. Conversely, facilities such as fire stations, schools, and stand-alone stations are notably less common.

*B. **Parallel Coordinates Plot for Compressed Natural Gas (CNG) information in 2021:***

Preprocessing of data:

1. Extraction of CNG Fueling Stations : Fueling stations with the CNG fuel type are extracted from the complete dataset. Subsequently, columns related to CNG are retained, while non-relevant columns are discarded.

2. Label Encoding for CNG Vehicle Class : The CNG vehicle class, being a categorical attribute, undergoes label encoding to convert it into numerical values, facilitating computational analysis.

3. Normalization of CNG Compression and Storage Capacities : Due to the wide range of values in CNG compression and storage capacities, a logarithmic transformation is applied to ensure uniformity and improve the robustness of subsequent analyses.

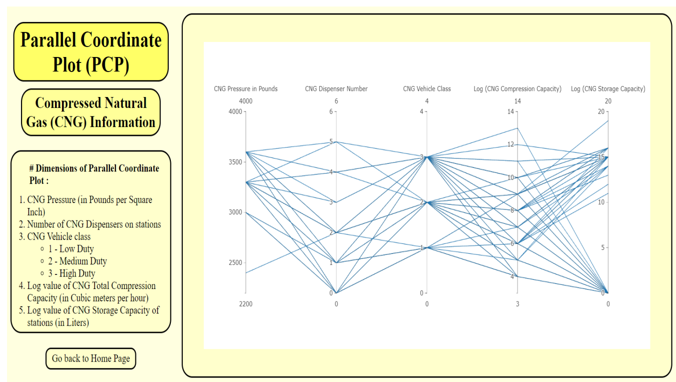

Fig. 2. Parallel Coordinates Plot for Compressed Natural Gas (CNG) information in 2021

The provided Parallel Coordinates Plot illustrates Compressed Natural Gas (CNG) information for the year 2021. The plot incorporates six dimensions, namely Fuel Type Code, Owner Type Code, Access Code, Facility Type, Latitude, and Longitude. Each of these dimensions contributes to the comprehensive visualization of CNG-related data, offering insights into the distribution and characteristics of fueling stations based on these key parameters. *Fig.2.* provides following observations and inferences :

Observations and Inferences:

1. CNG pressure of fueling stations : Most of the fueling stations have CNG pressure greater than 3000 pounds per square inch.

2. CNG storage capacity and CNG pressure : A positive correlation exists between the CNG storage capacity of fueling stations and CNG pressure; specifically, higher storage capacities correspond to elevated CNG pressures.

3. Vehicle Class and storage capacity : Generally, high and medium-duty vehicles tend to visit fueling stations with high storage capacities.

4. CNG compression capacity and storage capacity : There is a positive correlation between the CNG storage capacity and compression capacity of fueling stations. Specifically, higher storage capacities coincide with elevated CNG compression capacities.
5. Number of fuel dispensers : The number of fueling stations with fewer fuel dispensers exceeds those with a higher number of dispensers.
6. Storage Capacity : The majority of fueling stations exhibit CNG storage capacities within the logarithmic range of 12 to 16 liters.
7. Vehicle class and CNG Compression capacity : High duty vehicles require highly compressed CNG while medium and low duty vehicles uses less compressed CNG.

C. Parallel Coordinates Plot for Fuel Consumption information in 2021:

Preprocessing of data:
1. Transpose the Data : Convert the original data where years are columns and fuel types are rows to a new dataset where all fuel types and years become columns.
2. Exclude Low Consumption Fuel Types : Neglect fuel types such as RNG and Hydrogen due to their comparatively low fuel consumption. Consider including them after applying a logarithmic transformation to all columns to prevent data skewing.

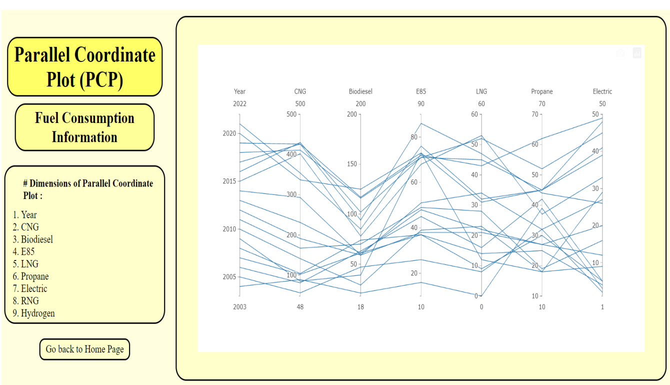

Fig. 3. Parallel Coordinates Plot for Fuel Consumption information in 2021

Embark on a visual exploration of fuel consumption trends across fueling stations from 2004 to 2021 through a Parallel Coordinates Plot. The axes—Year, CNG, Biodiesel, E85, LNG, Propane, and Electric—form an interconnected canvas, unraveling the dynamic shifts in fuel preferences. This visual narrative provides a concise yet comprehensive understanding of how different fuel types have evolved, acting as a valuable compass for informed insights into alternative fuel consumption dynamics. *Fig.3.* provides following observations and inferences :

Observations and Inferences:
1. CNG dominance : CNG consistently maintains the highest consumption from 2004 to 2021, with Biodiesel following as the next highest.
2. Year and Fuel consumption : A positive correlation is observed between the progression of years and the fuel consumption of nearly all fuel types, indicating a consistent upward trajectory in the consumption of alternative fuels over time.
3. CNG, Biodiesel and E85 : A linear relationship is evident among CNG, Biodiesel, and E85, signifying that an increase in one is associated with a corresponding increase in the others, and vice versa.
4. Electric and LNG :Electric and LNG exhibit comparatively lower consumption levels among all fuel types; however, their notable increases over the observed period are considerable.

IV. Conclusion

In conclusion, the Parallel Coordinates Plot has been instrumental in unraveling intricate patterns and correlations within alternative fueling stations. The exploration of key dimensions, such as fuel type, owner type, access, and location, has provided valuable insights into the spatial distribution and characteristics of alternative fuel infrastructure. This visualization not only enhances our understanding of the dataset but also serves as a powerful tool for sustainable energy planning.

The discerned patterns and correlations can inform strategic decision-making, facilitating the promotion and development of alternative fuel solutions. Overall, this report contributes to the broader discourse on sustainable energy by leveraging visual analytics to extract actionable intelligence from complex datasets.

References

[1] Heinrich, J., Weiskopf, D. (2013). State of the Art of Parallel Coordinates. Eurographics (state of the art reports), 95-116.
[2] McDonnell, K. T., Mueller, K. (2008, May). Illustrative parallel coordinates. In Computer Graphics Forum (Vol. 27, No. 3, pp. 1031-1038). Oxford, UK: Blackwell Publishing Ltd.

Colour Mapping of Sea Surface Temperature(SST) and Node Link diagram for American Football Infovis dataset

Bhavil Sharma - IMT2021041
IIIT-Bangalore
Data Visualization CS732

***Abstract*—This report features a Node Link diagram generated utilizing the NetworkX library, illustrating American Football team matchups. Specifically, the network focuses on ”American football games between Division IA colleges during the regular season of Fall 2000.” It is important to note that the dataset does not encompass results or home/away information. Additionally, a Color Mapping has been applied, employing a Jet color fill for the SciViz data spanning the duration of January to March 2021.**

I. Introduction

This report navigates the critical landscape of color mapping, a fundamental technique in data visualization that enhances interpretation by systematically assigning colors to data points. This method facilitates intuitive comprehension of intricate datasets, enabling pattern and trend identification. Concurrently, Node Link diagrams, powerful graphical representations of relationships within a network, contribute to understanding complex systems. With nodes representing entities and links illustrating interactions, this method is pivotal for unraveling intricacies. Together, Color Mapping and Node Link diagrams offer a robust means of visually exploring and comprehending complex datasets. This report provides a comprehensive exploration of color mapping, focusing on its application in the context of sea surface temperature. Furthermore, the Node Link diagram examines American football games between Division IA colleges during the regular season of Fall 2000.

II. My Work

In adherence to the specified guidelines, extensive experimentation was conducted concerning both color mapping and node-link diagrams. Ultimately, the most fitting visualizations have been selected for inclusion in this report. Additional visualizations and graphs are available in separate files, which have been archived and submitted in zip format for reference.

III. DataSets

A. SciViz - Jan - March 2021 — Sea Surface Temperature (SST):

Sea Surface Temperature (SST) is defined as the average temperature of the top few millimeters of the ocean. This temperature impacts the rate of all physical, chemical, and most biological processes occurring in the ocean. Sea Surface Temperature is globally monitored by sensors on satellites, buoys, ships, ocean reference stations, AUVs, and other technologies. Sea Surface Temperature monitoring tells us how the ocean and atmosphere interact, as well as providing fundamental data on the global climate system. This information also aids us in weather prediction. Sea Surface Temperature anomalies have been linked to shifting marine resources. With warming temperatures, we observe the poleward movements of fish and other species. Temperature extremes - both ocean heatwaves and cold spells - have been linked to coral bleaching as well as fishery and aquaculture mortality.

B. Infoviz - American Football:

In the context of Information Visualization (InfoViz), the dataset encompasses graph data in the form of an adjacency matrix representing the network structure. The data is formatted in whitespace-separated values, with each line denoting one edge of the network. Specifically, the network under consideration pertains to American football games between Division IA colleges during the regular season of Fall 2000. It is pertinent to note that the dataset intentionally omits information related to match outcomes as well as home/away designations.

IV. Node Link Diagram

A Node Link Diagram is a visual representation that illustrates the relationships and connections between entities, often referred to as ”nodes,” within a network. In this diagram, nodes represent individual entities, which can be people, objects, or any other discrete elements, and links depict the connections or relationships between them. The diagram is composed of nodes positioned in space, and lines (links) that connect nodes based on their relationships. Node Link Diagrams are commonly used in various fields such as network analysis, social sciences, biology, and information visualization to reveal patterns, clusters, and the overall structure of complex systems. They provide an intuitive and comprehensive way to understand the relationships and interactions present in a network, aiding in the exploration and analysis of intricate datasets.

A. Layout and Filtering

Following experimentation with multiple graph layouts, the Circular Layout was ultimately chosen for its efficacy in visually elucidating matchups between teams that have not played against each other, and conversely. Additionally, a topological filtering approach was implemented on the network based on degrees. This filtering method allows for the segregation of teams according to the number of matches they have participated in, enhancing the ability to discern and analyze team engagement within the network.

B. Descriptions

- Topological Filtering(Degree Filtering) : Nodes are filtered based on their degree, which represents the number of connections or links a node has. You might choose to display only nodes with a certain degree (e.g., nodes with high connectivity).
- Circular Layout: The Circular Layout is a graph visualization technique in which nodes and edges of a network are arranged in a circular or radial pattern. In this layout, nodes are positioned along the circumference of a circle, and the edges, representing connections between nodes, are drawn as arcs within the circle. The arrangement is such that nodes with stronger connections or similar attributes are often placed closer to each other, facilitating the identification of clusters or patterns within the network. The Circular Layout is particularly useful for visualizing symmetrical relationships, as it provides a clear and intuitive representation of connections, making it easier to observe patterns and relationships, especially in cases where cyclical or sequential patterns may be present in the data.
- NetworkX: NetworkX is a Python library designed for the creation, analysis, and visualization of complex networks and graphs. It provides a comprehensive set of tools and functions for working with nodes, edges, and their attributes, making it valuable for tasks such as network modeling, graph algorithms, and data visualization. NetworkX is widely used in various fields, including social network analysis, biology, and transportation planning, offering a flexible and efficient platform for exploring the structural and relational aspects of diverse networks.

C. Pros and Cons

Advantages:

Clarity of Relationships: Node Link Diagrams are effective in revealing relationships and connections between entities, providing a clear visual representation of the network's structure. Intuitive Understanding: The visual nature of Node Link Diagrams makes them intuitive and accessible for a broad audience, aiding in the quick comprehension of complex networks. Identification of Clusters: The diagram may facilitate the identification of clusters or groups within the network, helping to uncover patterns or substructures.

Disadvantages:

Visual Clutter: Depending on the size and complexity of the network, a Node Link Diagram can become visually cluttered, making it challenging to discern individual nodes and links. Edge Crossing: In dense networks, the crossing of edges can lead to confusion and hinder the accurate interpretation of relationships. Scale Sensitivity: The effectiveness of Node Link Diagrams can be influenced by the scale of the network. Extremely large networks may result in a diagram that is difficult to interpret.

V. Colour Mapping

Color mapping, also known as colormap or color scale, is a technique used in data visualization to represent numerical or categorical data using colors. In color mapping, each data value is associated with a specific color, creating a visual representation that allows viewers to interpret data patterns, variations, or trends more easily. Color maps are often applied to surfaces, graphs, or other visual elements to convey information in a way that is both intuitive and visually informative. Different color maps may be chosen based on the nature of the data and the goals of the visualization, and they play a crucial role in enhancing the interpretability of complex datasets.

A. Different Colour Maps

viridis, coolwarm, cividis, magma, and jet are names of different colormaps that can be used for visualizing data. Each colormap has its own characteristics, and they are designed for different types of data and visualization purposes.

The color maps were generated with respect to global minima and maxima. Subsequently, a GIF was created using the resulting color maps.

B. Descriptions

- viridis: A perceptually uniform colormap that is often recommended for displaying data. It is designed to be perceptually uniform, meaning that the change in color is perceived consistently across the range.
- coolwarm: A diverging colormap that transitions from cool colors (blues) to warm colors (reds). It is suitable for data where positive and negative deviations from a central value are of interest.
- cividis: Another perceptually uniform colormap, similar to viridis. It is designed to be accessible to those with color vision deficiencies.
- magma: A colormap with a perceptually uniform progression from black through various shades of red, yellow, and white. It is often used for highlighting details in data.
- jet: A popular colormap that was commonly used in the past, but it has some issues related to perceptual uniformity. It is not recommended for use in scientific visualizations due to artifacts and distortions in the perceived data.

C. Pros and Cons

Advantages:

Enhanced Interpretation: Color mapping enhances the interpretability of data by assigning different colors to distinct data values, making patterns and trends more apparent.

Visual Differentiation: It allows for clear differentiation between categories or ranges, aiding in the identification of variations and outliers in the dataset.

Intuitive Communication: Color is a powerful visual cue that can be easily interpreted, making color mapping an intuitive method for conveying information, especially in multidimensional datasets.

Facilitates Comparison: Color mapping allows for easy comparison between different data points or categories, enabling users to quickly grasp relative magnitudes.

Disadvantages:

Subject to Perception Differences: Individual perceptions of color can vary, leading to potential misinterpretations or inconsistencies in the understanding of data.

Colorblindness Concerns: Certain color maps may not be suitable for individuals with color vision deficiencies, limiting the accessibility of visualizations.

Limited to Screen Displays: The effectiveness of color mapping can be influenced by the medium of presentation; for example, printed materials may not reproduce colors accurately.

Potential for Misleading Impressions: Poorly chosen color scales or maps can distort the visual representation of data, leading to inaccurate or biased interpretations.

INFRENCES

D. Node Link Diagram:

1) Circular Layout:: In this visualization, the graph adopts a circular configuration, facilitating a clear discernment of unconnected nodes. The circular shape enhances the visual identification of disconnected nodes, contributing to an improved understanding of the network structure.

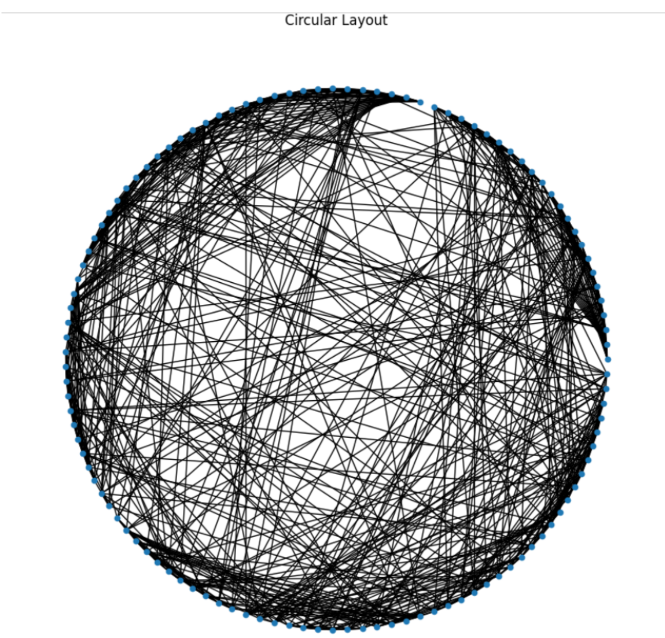

Fig. 1. Circular Layout of matches played by American Football teams

Here, The diagram facilitates the identification of clusters or groups within the network, helping to uncover patterns or substructures. This way, we can easily find out which cluster of teams play together often and which don't. And which teams have never played each other.

2) Degree Filtering:: In this context, teams are undergoing a filtration process based on the quantity of games they have participated in.

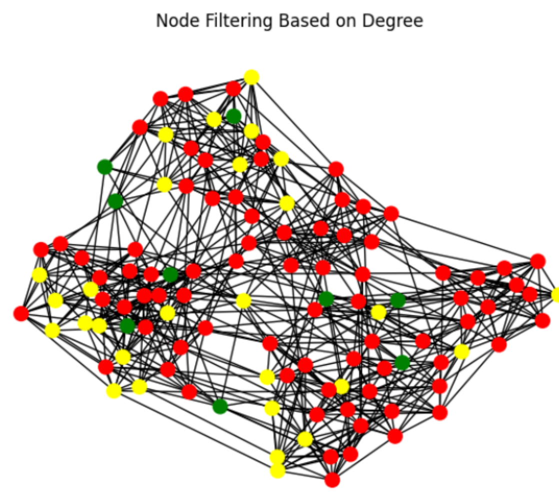

Fig. 2. Degree Filtering of American Football teams

In this representation, teams highlighted in red exhibit the highest level of experience, denoting a greater number of matches played with various teams. Those in yellow signify a mid-tier experience, while the green teams indicate a novice status in the game. This categorization enables a discerning analysis of the experience levels within the network for each respective team.

E. Colour Mapping:

1) January 2021: Here is the colour mapping on 2nd Jan 2021

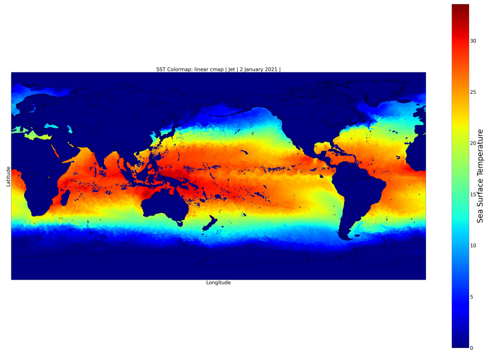

Fig. 3. Colour map on January 2 for Sea Surface Temperature

2) February 2021: Here is the colour mapping on 8th Feb 2021

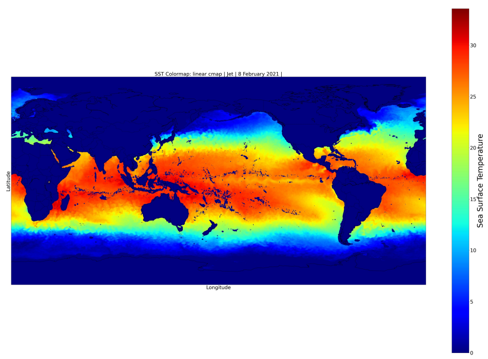

Fig. 4. Colour map on February 8 for Sea Surface Temperature

3) March 2021: Here is the colour mapping on 5th March 2021

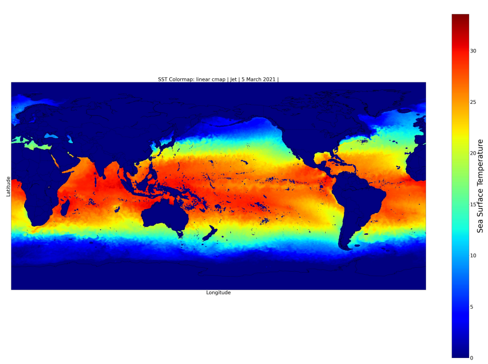

Fig. 5. Colour map on March 5 for Sea Surface Temperature

Through the examination of these three representations, it becomes evident that the surface temperature exhibits a consistent pattern within a given region. While the boundaries may undergo minor fluctuations, the overall climatic conditions persist throughout the designated three-month period from January to March 2021. Notably, the Sea Surface Temperature (SST) remains elevated in proximity to the equator and gradually decreases as one moves towards the poles. The visual depiction resembles the tropical lines (Tropic of Cancer and Tropic of Capricorn), delineating regions of varying SST (as indicated by the yellow region).