DV Assignment-2

1st Aryan Vaghasiya
IMT2022046
aryan.vaghasiya@iiitb.ac.in
Data Titans

2nd Areen Vaghasiya
IMT2022048
areen.vaghasiya@iiitb.ac.in
Data Titans

3rd Rutul Patel
IMT202202
rutul.patel@iiitb.ac.in
Data Titans

***Abstract*—This document is a model and instructions for LaTeX. This and the IEEEtran.cls file define the components of your paper [title, text, heads, etc.]. *CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

For A2, the authors *Aryan et al*. have implemented SciVis methods and InfoVis methods on two different datasets.
For SciVis:
gridMET link taken from [1] is a dataset of daily high-spatial resolution (4-km, 1/24th degree) surface meteorological data covering the contiguous US from 1979-yesterday(40 years). We have also extended these data to cover southern British Columbia in our real time products. By effectively utilizing scivis techniques, the authors *Aryan et al*. have tried to gain a deeper understanding of climate patterns, make informed decisions, and mitigate the impacts of climate change. These data can provide important inputs for ecological, agricultural, and hydrological models.

For InfoVis:
For this the authors have chosen the dataset jazzData **Jazz musicians network: List of edges of the network of Jazz musicians** used for network visualizations such as Node-link diagrams/ Treemaps/Parallel coordinate plots. By visually representing complex relationships between musicians, genres, and historical periods, the authors *Aryan et al.* have tried to gain a deeper understanding of the jazz music landscape.

II. Scientific Visualization

A. DataSet

*1) For the **SciVis** Dataset from **gridMET**::* **Primary Variables:**
Maximum temperature, minimum temperature, precipitation accumulation, downward surface shortwave radiation, wind-velocity, humidity (maximum and minimum relative humidity and specific humidity. **Derived Variables:**

- tmmx - Maximum temperature (K)
- tmmn - Mininum temperature (K)
- pr - Precipitation Amount (mm)
- rmax - Maximum relative humidity (%)
- rmin - Minimum relative humidity (%)
- sph - specific humidity (mass fraction)

TABLE I
Derived Variables and Their Dependencies on Primary Climate Variables

Derived Variable	Primary Variables
Reference Evapotranspiration	Max/min temp, solar_rad, humidity, wind speed
Energy Release Component	Temperature, humidity, precipitation
Burning Index	ERC, wind speed
Dead Fuel Moisture (100/1000 hr)	Temperature, humidity, precipitation
Mean Vapor Pressure Deficit	Temperature, humidity
Palmer Drought Severity Index	Temperature, precipitation over time

- srad - Surface downward shortwave radiation(W/m^2)
- th - wind direction $(\deg)$
- vs - wind speed (m/s)
- erc - Energy Release Component (NFDRS fire danger index)
- bi = Burning Index (NFDRS fire danger index)
- fm100 - 100-hour dead fuel moisture (%)
- fm1000 - 1000-hour dead fuel moisture (%)
- vpd - vapour pressure deficit (kPa)
- etr - evapotranspiration reference (mm)

*2) The **Jazz**-dataset used for **InfoVis**:* This dataset consists of nodes and edges for the social network of the jazz musicians.

B. Color Maps

Color mapping, also known as a colormap or color scale, is a technique in scalar field visualization used to represent variations in a scalar quantity across a spatial domain. In this context, a scalar field is a function that assigns a scalar value to each point in space, representing physical properties like temperature, pressure, density, or other scalar attributes. By assigning colors to specific scalar values, color mapping visually communicates the variations within the scalar field, making it easier to interpret patterns and distributions of the quantity across the domain. Choosing appropriate colors and their arrangement is essential for effectively conveying information about the scalar field.

Here the dataset used is of gridMET dataset
It has been implemented using matplotlib and other libraries.

Experiments On Color Mapping Palettes: Various color palettes were compared, such as PiYG, PrGN, BWR, which

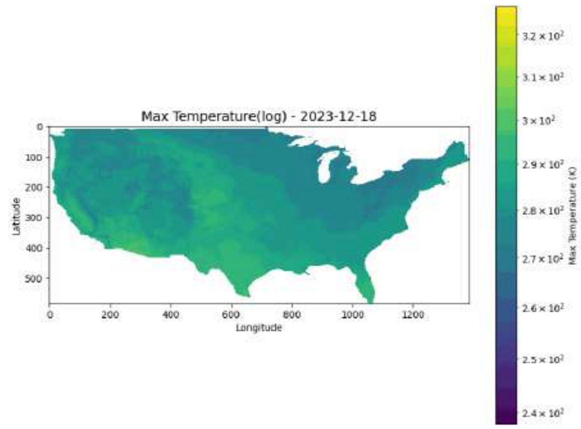


Fig. 1. viridis-log

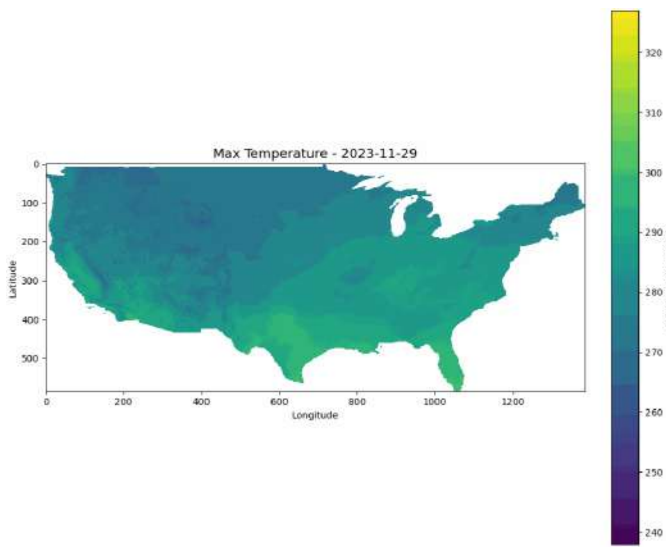


Fig. 2. viridis-continuous

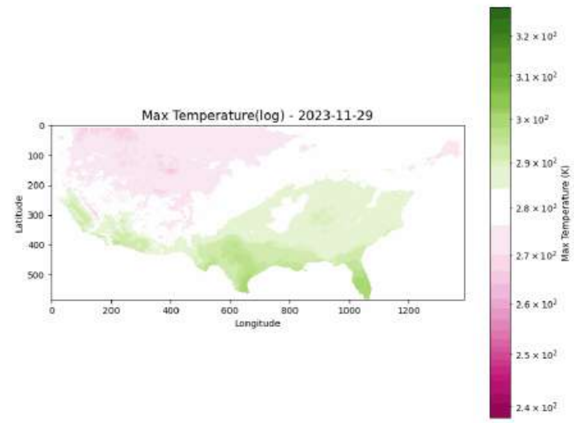


Fig. 3. PiYG - log

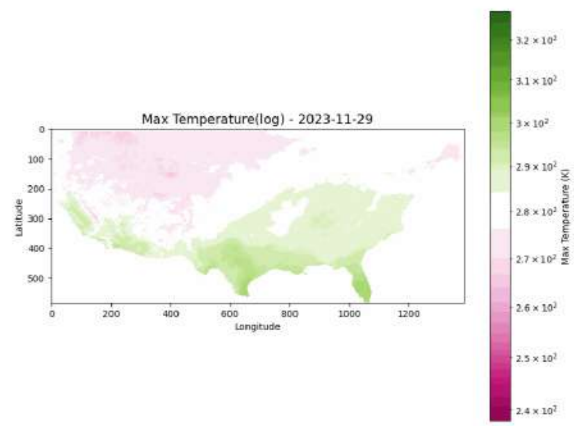


Fig. 4. PiYG - contiuous

are diverging color palettes, and Hot, Magma, Viridis, which are sequential color palettes. We also observe that the majority of scalar values in our dataset are lessthan the center of the range of values, so a convergent color palette is better for visualization.

These observations led to the choice of proceeding with a sequential color palette which does not introduce artefacts. For our purpose, we chose to proceed with the sequential Viridis color palette, which has the added benefit of being a perceptually ordered color map. Henceforth, we shall consider only this color map for further experiments and visualizations.

Experiments On Parametric Mapping: Primarily, there are two types of parametric color mapping: using global maxima and minima, and using local maxima and minima.

Global Mapping is suitable when one wants a consistent color scheme for the entire dataset, making it easier to observe long-term trends and identify outliers across all timesteps. It is useful when we want to emphasize the overall distribution of scalar values. This approach ensures that color representations remain stable and con sistent, allowing viewers to compare different timesteps directly based on the global scalar value range.

Local Mapping is beneficial when you want to focus on temporal changes or localized features in the data. It provides a more granular view of each timestep, which can be helpful in tracking specific events or phenomena as they evolve over time. Using local maxima and minima can help highlight temporal variations and anomalies in the scalar field over time, but it may make direct comparisons between timesteps

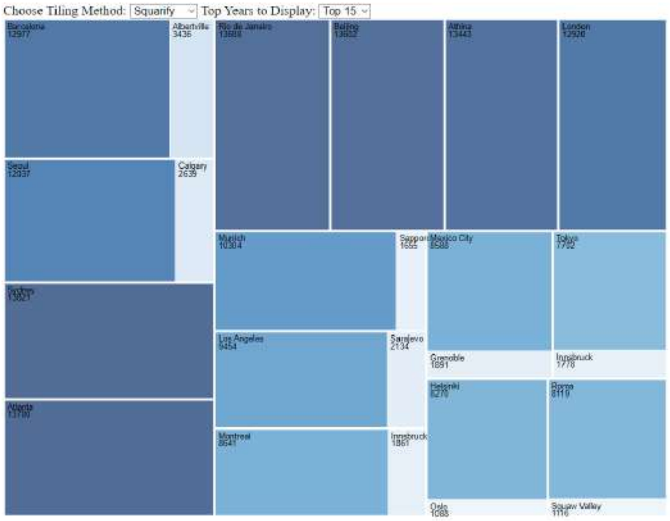


Fig. 5. Event Participation by Year and City Squarify method

more challenging.

Experiments On Color Mapping Scales: Three types of scales, namely Continuous, Discrete, and Logarithmic, were considered for mapping scalar field values to colors.

- **Logarithmic Scale**: As the logarithm of 0 (the global minima) approaches negative infinity, we consider 10^2 as the minimum when using a logarithmic mapping scale. The clip parameter is set to True while creating the normalization object, ensuring that values below 10^2 are also mapped to the lowest color on the color map. Logarithmic scales are useful when the scalar values are very large.
- **Continuous Scale**: In order to use a continuous scale, we first normalize the data. When the scalar field was visualized without normalization, no discernible differences were observed. Thus, we decided to visualize the field without normalization, as can be seen in Fig 2. Note that to make the visualization easy to interpret, the color bar is representative of the actual scalar value.
- **Discrete Scale**: We first linearly scale the values using the Normalize class in matplotlib.colors, and then introduce discrete color bands of length 0.2 each using the Bound aryNorm class. These are mapped to specific colors in the Viridis color palette. We also use a special color, red, to indicate values values which are notably high (more than 2.0 on the new scale).

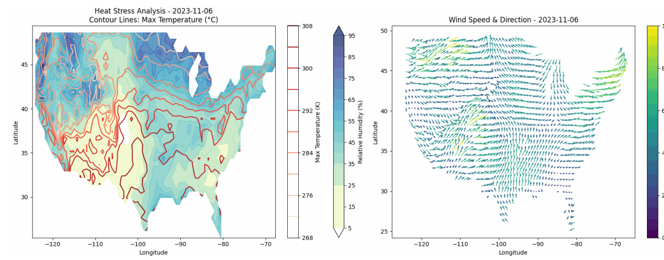
C. *gridMET SciVis - Contour Maps*

There are a number of parameters that we can use to analysis the meteorological data of gridMET. Some of them which can be visualized using contour maps that the authors Aryan et al. have used here are :

- Heat Stress Analysis
- Flood Potential Analysis
- Agricultural Stress analysis
- Fire Danger Analysis
- EvapoTranspiration Stress Analysis
- Cold Wave Detection Analysis

D. *Heat Stress Analysis*

Here's a detailed breakdown of the reasons for choosing the parameters, time period, and the suitability of this visualization for analyzing heat stress, along with the inferences that can be drawn:



Choice of Parameters

- **Maximum Temperature**:
  By using maximum temperature, we capture the peak heat levels in a day, which have the most significant impact on heat stress.
- **Minimum Relative Humidity**:
  Lower humidity levels make hot temperatures feel even hotter due to reduced evaporative cooling, which is essential for human bodies to regulate heat. Low relative humidity also increases the risk of dehydration, making it a critical component in assessing heat stress.
- **Wind Velocity and Direction**:
  Wind plays a dual role in heat stress analysis. Gentle breezes can provide relief by enhancing cooling through convection, while strong, dry winds (e.g., Santa Ana winds) can amplify heat stress.

Time Period Selection (November to January)

- **Seasonal Variation and Risk**: It was chosen keeping urban heat island effect in mind. Although heat stress is often a summer issue, the November to January period can still pose risks in regions experiencing abnormal warming, particularly in urban areas as less greenery and more concrete are prone to heat retention.

Inference and Analysis

- **Contour plot** was used for quick identification of regions with high temperature and low humidity as well as for highlighting gradient changes to indicate critical risk areas.
- **QuiverPlot** for wind captures both the magnitude and orientation of wind, crucial for assessing whether winds might relieve or exacerbate heat. The combined analysis of both gives a clear idea of heat stress patterns.
- **Identification of High-Risk Areas**: Areas where high temperatures and low humidity coincide can be quickly

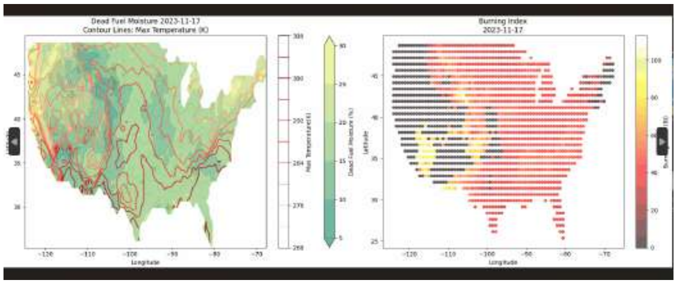


Fig. 6. FireDanger

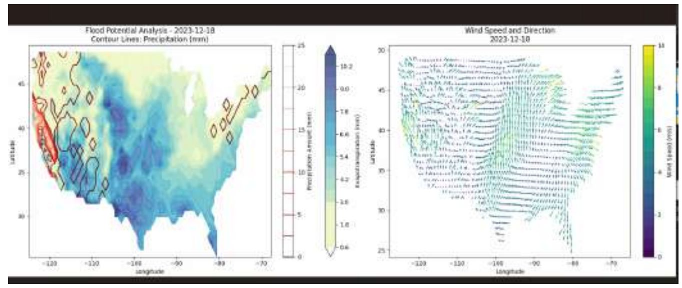


Fig. 7. Flood Potential

identified as high-risk zones for heat stress. This is particularly relevant in urban settings with dense infrastructure, which contributes to the Urban Heat Island effect.

- **Wind's Role in Heat Amplification or Relief**: The wind analysis reveals how different wind patterns impact heat stress. For instance, high-speed winds from specific directions (dry, hot winds) may worsen conditions in affected areas, while gentle breezes might offer some cooling.
- **Daily Variability and Consistency of Heat Stress**: Observing conditions over multiple days helps determine if certain areas consistently experience higher stress levels, aiding in targeted resource allocation and policy planning.

E. Fire Danger Analysis

Fire danger analysis mentions how the statistics vary that may cause an outbreak of fire. Here also a total of 3 parameters are used for analysis. **Choice of Parameters**

- **Maximum Temperature**:
High temperatures are a primary factor in increasing wildfire risk. Warmer temperatures dry out vegetation, reducing moisture content and making it more susceptible to ignition.
- **Dead Fuel Moisture (100-hour and 1000-hour)**:
This variable represents moisture content in dead vegetation, broken down into medium- to long-term categories based on how long it takes for the vegetation to respond to changing environmental conditions. Low dead fuel moisture levels indicate dry vegetation, which is more flammable.
- **Burning Index**:
This index measures fire intensity by combining temperature, humidity, and fuel moisture levels. A high burning index value means conditions are favorable for intense fires.

Reasons for Visualization Choices

- **Contour Lines for Maximum Temperature**: Contours show temperature gradients across the area, making it easier to spot zones with higher temperatures that are more vulnerable to fire outbreaks.
- **Filled Contours for Dead Fuel Moisture**: Filled contours visually indicate areas with different moisture levels. Areas with low dead fuel moisture are highlighted, showing where vegetation is most flammable.
- **Scatter Plot for Burning Index**: The burning index is shown with color-coded markers that visually emphasize fire intensity. This helps identify high-risk zones where fire behavior could be intense.
-

Inferences and Detailed Analysis

- **High Temperature and Low Dead Fuel Moisture**:
By visualizing these variables together, areas with high temperatures and low dead fuel moisture become immediately noticeable. Such areas indicate a "tinderbox" effect, where vegetation is dry and highly flammable, creating ideal conditions for wildfire ignition and spread.
- **Identification of Fire-Intense Areas**: The burning index provides a direct measure of potential fire intensity. Higher burning index values in regions with high temperatures and low moisture levels suggest zones where wildfires, if ignited, would likely be severe and difficult to control.
- **Influence of Seasonal Changes on Fire Risk**: By using data across multiple days, this visualization can highlight how seasonal changes impact wildfire risk. For example, conditions in late summer and fall, which tend to be hotter and drier, might show higher burning index values and lower fuel moisture levels than in other seasons.
- **Potential Impact on Firefighting and Resource Allocation**: By analyzing burning index hot spots, fire management teams can focus resources on high-risk areas, preparing firefighting efforts, and planning resource allocation based on predicted fire behavior. Areas with high burning index and low moisture might be designated for higher fire surveillance and faster response times.
- **Early Warning for At-Risk Communities**: For urban and rural areas close to wildland regions, the visualization serves as an early warning. High temperature and low moisture levels near populated regions indicate the need for preventive measures like creating defensible space around properties.

F. *Flood Potential Analysis*

Choice of Parameters

- **Precipitation Accumulation** (P)
  High precipitation directly contributes to increased water levels in soil, rivers, and streams, making it the primary factor in assessing flood potential.
- **Evapotranspiration Rate (ETR)**
  ETR represents the rate at which water is removed from the soil and vegetation. Low ETR indicates reduced moisture loss, which contributes to soil saturation, increasing flood risk.
- **Wind Speed** (W)
  Wind speed affects ETR by influencing how quickly moisture evaporates from the soil and plants. Low wind speed reduces evaporation, leading to higher soil moisture levels and potentially increasing flood potential.

Visualization Approach

The following visualizations are used to illustrate flood potential based on these variables:

- **Contour Lines:** Represent **Precipitation Accumulation** levels across regions, helping to identify areas of high rainfall and runoff.
- **Filled Contours:** Depict **Evapotranspiration Rate (ETR)**, with lower ETR values highlighting regions at higher risk of water saturation.
- **Quiver Plot:** Displays **Wind Speed** vectors, showing the speed and direction of wind flow that influences ETR and, by extension, soil moisture.

Inference and Analysis

Analyzing the data for flood potential provides the following insights:

1) **High Precipitation with Low ETR and Low Wind Speed**: Regions with heavy precipitation, low ETR, and minimal wind speed represent areas with high flood potential. Heavy rainfall raises soil water levels, while low ETR and wind speed mean that moisture is not being effectively removed from the soil, increasing saturation and runoff.
2) **Evapotranspiration and Wind Impact on Flood Potential**: In areas where ETR is high and wind speed is moderate or high, moisture is more likely to evaporate, even if precipitation is significant. These regions have a lower flood risk due to effective moisture removal. This insight is valuable for differentiating between flood-prone and resilient areas.
3) **Flood Potential in Wind-Protected Valleys or Urban Areas**: In regions such as valleys or urban areas where wind speeds are naturally low, soil moisture tends to accumulate. Heavy rainfall in these areas poses a higher flood risk since low wind means reduced moisture evaporation. This analysis can guide flood management strategies in topographically vulnerable areas.
4) **Temporal Flood Risk Analysis:** By visualizing these variables over multiple days, one can assess flood risk trends over time. For instance, consecutive days of heavy precipitation with low ETR and wind speeds may indicate a cumulative increase in flood risk as soil becomes more saturated.
5) **Implications for Agricultural and Urban Planning:** High precipitation and low ETR in agricultural areas may lead to soil erosion, crop loss, or even localized flooding. In urban planning, identifying flood-prone areas helps in developing better drainage systems, green spaces for water absorption, and infrastructure that can handle increased water loads.

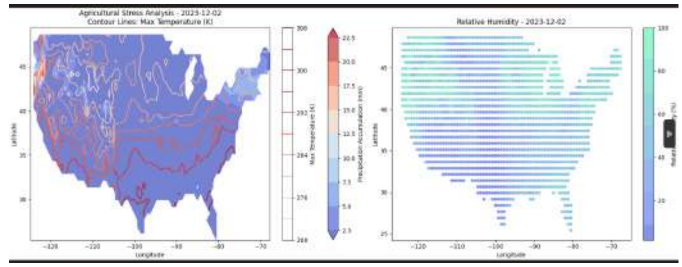


Fig. 8. Agri

G. *Agricultural Stress analysis*

This visualization approach allows agricultural managers and farmers to see how different environmental factors combine to affect crop health, helping them take preemptive actions like scheduling irrigation, adjusting planting strategies, or choosing crops suited to the current conditions.

Reasons for Choosing the Parameters

- Maximum Temperature: High temperatures increase water evaporation from the soil and increase plant transpiration, creating stress conditions for crops. Prolonged exposure to elevated temperatures without sufficient water supply can cause heat stress in crops, leading to reduced growth and yield.
- Precipitation Accumulation: Rainfall is essential for crop hydration, and low precipitation levels lead to water scarcity. Monitoring rainfall accumulation helps assess areas experiencing drought or low water availability, crucial for understanding stress conditions in agriculture.
- Relative Humidity: Humidity affects evapotranspiration rates, which influence the water available to crops. Low relative humidity increases water loss from the soil and plants, leading to faster soil drying and exacerbating drought conditions. Monitoring humidity levels, especially in combination with temperature, provides a clear picture of atmospheric conditions that could negatively impact crops.

Inference and Analysis:

- High Temperature and Low Humidity: Regions experiencing high temperatures and low humidity are at risk of severe agricultural stress. These conditions increase

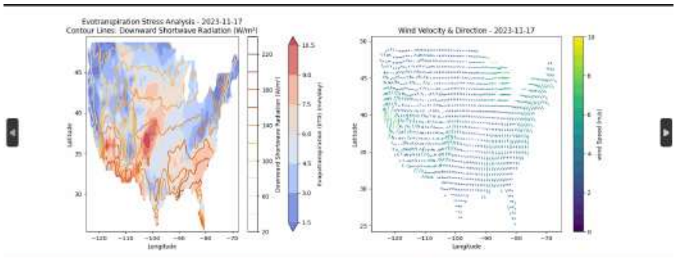


Fig. 9. evapotranspiration

evapotranspiration, leading to rapid soil moisture loss and stressing crops, which can reduce yields.

- Low Precipitation: Areas with minimal rainfall and low humidity create an even more challenging environment for crops, as soil moisture quickly depletes. Such regions are high-priority for irrigation to prevent crop damage.
- Targeted Agricultural Intervention: This visualization aids in pinpointing at-risk agricultural areas, allowing farmers and policymakers to address issues like water scarcity and crop heat stress more effectively.
- Pest and Disease Vulnerability: Stress conditions (e.g., high temperature, low humidity) can make crops more susceptible to pests and diseases. Dry conditions may stress plants, weakening their natural defenses and making them more prone to infestations or diseases that thrive in dry, warm climates. This visualization could help predict areas where pest control efforts might be needed.
- Impact on Crop Types: The analysis can inform crop selection, suggesting that more drought-resistant or heat-tolerant crop varieties may be better suited for areas with high temperatures and low precipitation. By identifying regions prone to agricultural stress, policymakers and farmers can consider planting crops that are more resilient to adverse conditions.

H. EvapoTranspiration Stress

High solar radiation and wind speed can increase evapotranspiration, impacting water demand for plants and agriculture. Areas with high ET0 might require increased irrigation.

Reason for parameters

- Downward Surface Shortwave Radiation: Shortwave radiation is the primary energy source that drives the evaporation and transpiration processes. High solar radiation can increase the potential for evapotranspiration (ET), particularly in regions with sufficient moisture in the soil and vegetation. This makes it an essential parameter for understanding water stress in agricultural regions or ecosystems.
- Reference Evapotranspiration (ET): ETr represents the amount of water evapotranspired by a reference crop under optimal conditions. It is directly influenced by factors like temperature, radiation, wind, and humidity. Monitoring ETr helps in understanding the water needs for irrigation, especially in arid or semi-arid regions where crops might require supplemental watering.
- Wind Velocity: Wind can enhance evapotranspiration by increasing the rate of evaporation from soil surfaces and transpiration from plants. Strong winds can increase water loss, especially in regions with high temperatures and limited moisture. Wind direction and speed are crucial for understanding localized evapotranspiration patterns and water demand.

Inferences and Analysis

The period between November and January represents the winter season in the northern hemisphere. During this time, evapotranspiration rates tend to be lower compared to the summer months, but this can vary depending on location .

- Increased Evapotranspiration in High Radiation Areas: Regions with high downward shortwave radiation and high ET0 will experience increased water demand, making them vulnerable to water stress. These areas may require more efficient irrigation systems or water conservation measures to maintain crop health.
- Wind-Enhanced Evapotranspiration: Strong winds, especially in regions with high radiation, can significantly accelerate evapotranspiration. The wind speed vectors provide a dynamic view of areas that are prone to higher water loss. This is especially important for dryland farming areas where water conservation and efficient irrigation systems are critical.
- Identification of Critical Zones for Irrigation: By combining ET0, shortwave radiation, and wind velocity, the visualization can help identify areas where evapotranspiration is highest and where crop irrigation requirements might be the most urgent. Areas with both high solar radiation and strong winds could experience rapid soil moisture depletion, and timely irrigation planning can prevent crop stress.
- Impact of Wind on Water Conservation: Wind can exacerbate water loss from soil surfaces. Regions with high wind velocity could suffer from more rapid soil drying, even if they have relatively lower radiation. The wind vectors allow us to pinpoint such areas and plan for windbreaks, mulching, or soil moisture management strategies.

I. Cold Wave Detection

Extremely low temperatures combined with high wind speeds and low humidity are indicative of cold waves. This is useful for cold-weather alerts in populated or vulnerable areas.

Reason for Parameters

- Minimum Temperature: Cold waves are characterized by extremely low temperatures. Monitoring minimum temperature is crucial for detecting the onset of such extreme weather events. Regions experiencing unusually low minimum temperatures can indicate the development

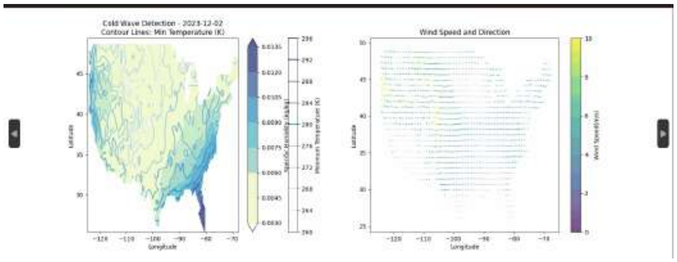


Fig. 10. coldwave

of cold waves, which may pose risks to both human health and infrastructure.

- Wind Speed: Wind speed plays a significant role in enhancing the effects of cold temperatures, especially during cold waves. High wind speeds increase the rate of heat loss from the body through convection (wind chill), making cold weather feel even colder. Strong winds can exacerbate the negative impacts of low temperatures, particularly in populated or vulnerable areas.
- Specific Humidity: Specific humidity measures the amount of moisture in the air. Low humidity levels are often associated with cold, dry air, which is common during cold waves. Low moisture in the atmosphere can further intensify the feeling of coldness (through the wind-chill effect) and contribute to harsh environmental conditions. Monitoring this parameter helps in understanding how the air's moisture content contributes to the overall cold wave severity.

Inferences and analysis

- Cold Wave Identification: The combination of low minimum temperatures and high wind speeds in the visualization can immediately identify areas where cold waves are likely to occur. The wind chill effect, especially in areas where wind speeds are high and temperatures are low, is clearly shown, indicating where the cold wave's impact will be most severe.
- Wind-Enhanced Cold Effects: The quiver plot showing wind direction and speed can help identify where the wind is most likely to exacerbate cold conditions. For example, in regions where cold winds are blowing over already low-temperature areas, the combined effect will lead to greater discomfort and danger to human health. The wind vectors also show the direction of the cold air movement, which can aid in forecasting the progression of the cold wave.
- Low Humidity and Dry Cold: Low specific humidity values in areas with low temperatures indicate a dry cold. This condition is typical of cold waves, where low moisture in the air contributes to the increased severity of the cold. The filled contour plot for humidity will show which areas experience the driest conditions during the cold wave, which can be important for understanding the potential health risks (e.g., dehydration, frostbite).

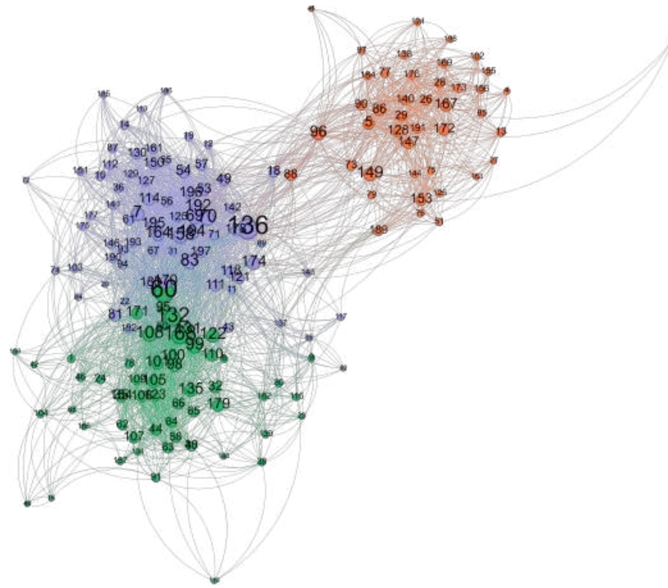


Fig. 11. Yifan Hu Proportional

- Health and Safety Risks: By visualizing the combination of low temperatures, wind speed, and humidity, the analysis can highlight regions where cold waves pose the greatest risk to human health and safety. For example, areas with high wind speeds and low temperatures may experience significant wind chill, increasing the risk of frostbite, hypothermia, and other cold-related illnesses. This helps to prioritize which areas need immediate attention in terms of health advisories, shelter, and heating resources.

INFOMATION VISUALIZATION

J. NodeLink Diagrams

To thoroughly address which layout algorithm is best for visualizing a jazz musician social network, let's break down each of the algorithms—Yifan Hu, Fruchterman-Reingold, and ForceAtlas2—in detail, considering aspects like scalability, efficiency, and suitability for social network data. Best For: Large but relatively sparse networks where proportional distance between nodes is important.

Yifan Hu Proportional Layout

Yifan Hu's algorithm is a force-directed method that introduces a hierarchical or multilevel approach to optimize the placement of nodes. The algorithm reduces complexity by iteratively collapsing nodes into coarser levels, positioning nodes at each level, and then expanding back to finer levels.
Force Model: Like other force-directed layouts, it uses attractive and repulsive forces between nodes to position them in such a way that the graph becomes easier to interpret visually. However, Yifan Hu introduces additional proportional scaling, ensuring that nodes are spaced proportionally to the distances between them.

- Advantages:

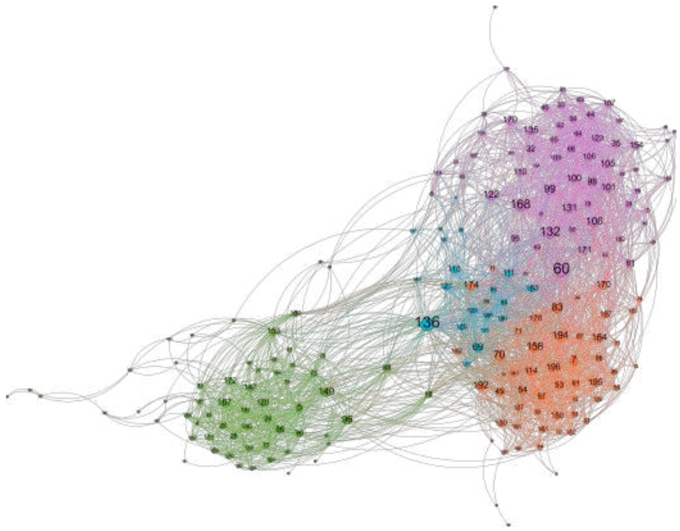


Fig. 12. ForceAtlas2

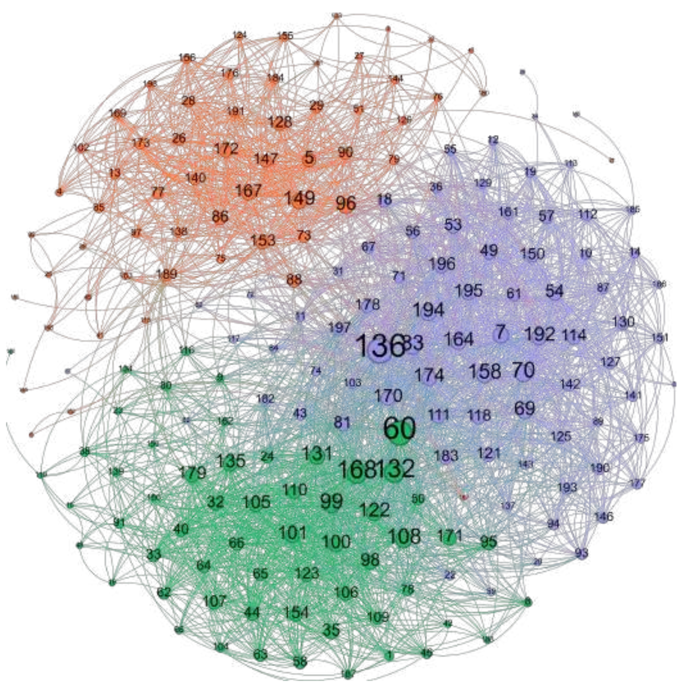


Fig. 13. FR

  - Scalability: The multilevel approach makes it highly scalable. It can handle larger datasets more effectively than traditional force-directed algorithms like Fruchterman-Reingold.
  - Balanced Distribution: Yifan Hu focuses on distributing nodes evenly, which often results in clearer representations, especially when dealing with large, sparse graphs.
- Disadvantages:
  - Cluster Separation: It may not cluster nodes as tightly or intuitively in dense networks, which might be a drawback for social networks where clustering and community detection are important.
  - Processing Time: Although scalable, it may still take longer to compute compared to newer methods like ForceAtlas2 for very dense graphs.

Fruchterman-Reingold (FR) Layout

FR is one of the oldest and most widely used force-directed algorithms. It models the graph as a physical system where nodes are charged particles repelling each other, while edges act as springs pulling nodes together.

Attractive forces are proportional to the square of the distance between connected nodes, and repulsive forces are inversely proportional to the distance between any two nodes.Best For: Small to medium-sized networks, where relationships between nodes are not overly complex or densely connected.

advantages:

- Intuitive Visualization: The FR layout often produces visually intuitive results, with related nodes positioned near each other and unrelated nodes pushed far apart.
- Simplicity: FR is relatively easy to implement and understand, making it a popular choice for smaller networks.

Disadvantages:

- Convergence: For large networks, it may take a long time to converge to a stable layout, and the result may not be optimal for high-density graphs.
- Computational Complexity: As the number of nodes increases, the FR layout becomes computationally expensive, especially for large, dense graphs like social networks.

ForceAtlas2 Layout

ForceAtlas2 is an optimized version of force-directed algorithms specifically designed for social networks. It introduces several enhancements over the traditional force-directed approaches, including improved performance on large graphs and better clustering for highly connected components.

Force Model: Like FR, it uses attractive and repulsive forces but is optimized to handle large-scale graphs. It simplifies repulsive force calculations, allowing the layout to run faster even on larger networks. It also offers customizable parameters to control attraction and repulsion between nodes, adjust the strength of forces, and handle edge weights. Best For: Large, dense networks with many communities, such as a social network of jazz musicians where collaboration patterns may form distinct clusters.

advantages:

- Speed and Scalability: ForceAtlas2 is highly efficient and can handle very large datasets, such as complex social networks with thousands of nodes and edges. Its speed comes from a combination of multithreading and force approximations, making it faster than traditional methods like FR.
- Clustering and Community Detection: One of the standout features of ForceAtlas2 is its ability to clearly display clusters or communities within the network. This makes it especially well-suited for social networks like those

involving jazz musicians, where distinct groups (e.g., based on genre, location, or collaboration history) are likely to emerge.

- Adjustability: It allows for fine-tuning parameters such as gravity (which controls how tightly or loosely clusters are formed), scaling (to prevent node overlap), and speed. This makes it highly customizable for specific network properties.

disadvantages

- Visual Overlap: If not configured correctly, ForceAtlas2 can sometimes cause visual overlap in dense parts of the network, though this can be mitigated through parameter adjustment.
- Potential for Over-Clustering: In some cases, nodes may cluster too tightly, obscuring individual relationships within larger clusters.

Conclusion: Best for Jazz Musician Social Network

- ForceAtlas2 is likely the best choice for visualizing a jazz musician social network because:
  The algorithm's efficiency in handling large datasets ensures that even if the network is extensive (e.g., including thousands of musicians), it will still perform well.
- Jazz musician networks are typically dense, with many collaborations forming clusters based on location, genre, or even recording sessions.
- The flexibility of ForceAtlas2 allows for fine-tuning, ensuring that the visualization can be adjusted to emphasize or de-emphasize certain relationships, making it a highly customizable solution.
- ForceAtlas2 is specifically designed to handle this kind of dense network, and its ability to clearly display communities will make it easier to interpret clusters of musicians who frequently collaborate.

Yifan Hu could be a good alternative if a proportional layout is preferred, but it might not handle dense clusters as well as ForceAtlas2.

Fruchterman-Reingold, while useful for smaller datasets, would likely struggle with the complexity and density of a large jazz musician network.

K. TreeMaps

1) Data: The dataset used is from of Jazz.net. It's a dataset of Jazz musicians network: List of edges of the network of Jazz musicians. P.Gleiser and L. Danon , Adv. Complex Syst.6, 565 (2003). The graph is unweighted and undirected.
   There are 198 nodes and 2742 edges connecting them.
2) Visualization: The visualization is done using Gephi ref by implementing different layouts.
3) The Fruchterman Reingold Algorithm:

1) **Description:** A treemap is a space-efficient visualization method for hierarchical data, representing nested categories as rectangles whose sizes correspond to data values. In implementation, a treemap begins by structuring data hierarchically, often using a root node with nested child nodes for each category or subcategory. Each node's value determines the area of its corresponding rectangle, which allows viewers to quickly compare relative sizes within the entire dataset. Implementing a treemap in JavaScript commonly involves using the D3.js library num, which provides tools to create hierarchies and layouts, such as numd3.hierarchy for structuring data and numd3.treemap for calculating the positions and dimensions of each rectangle.
2) **Data:** The dataset used is num'120 years of Olympic data'. The dataset contains visually perceptive hierarchical and non-hierarchical data that can be visualized using Treemaps.
   A number of pre-processing had to be done to handle the null values in the 'medal' column of the dataset which has been used extensively in this visualization.
3) Interactions:
   - **Tiling Method Selection:** A dropdown menu labeled "Choose Tiling Method" allows users to select different tiling algorithms (e.g., Squarify, Slice, Dice, Slice-Dice). Each option rearranges the treemap layout, changing how the rectangles are partitioned based on the selected algorithm.
   - **Top Countries Filter:** Users can filter by top countries through a dropdown, focusing on the most significant categories (e.g., top medal counts). Users can choose from various predefined values (e.g., Top 5, Top 10, Top 15), which dynamically updates the treemap, focusing on the most relevant categories and allowing for a more detailed examination of key data points.
   - **Hover Tooltip:** Tooltips appear on hover, displaying hierarchical paths (e.g., Country vs Sport) and values, providing quick insights without cluttering the view.
   - **Hierarchy Distinction:** Country-level nodes are styled with darker borders, visually differentiating primary categories from subcategories for easier reading.

   In this treemap implementation, different tiling algorithms provide flexibility in how rectangles are arranged within the hierarchy, each offering a unique approach to representing data:

   - **Squarify:** Balances rectangle aspect ratios to keep shapes close to squares, making it easy to compare areas visually.
   - **Slice:** Arranges data in horizontal rows, ideal for a linear view of hierarchical data.
   - **Dice:** Lays out rectangles in vertical columns, complementing Slice with a contrasting vertical

organization.

- **Slice-Dice:** Alternates between horizontal and vertical layouts, balancing shape and orientation for multi-level hierarchies.

4) **Visualization 1: Event Participation by Year and City:** The Event Participation Treemap provides a visual breakdown of the number of athletes participating in events by Year and City, allowing users to explore which cities hosted more events or attracted larger numbers of athletes.
The treemap's color gradient reflects the number of participants, with darker shades indicating higher counts.

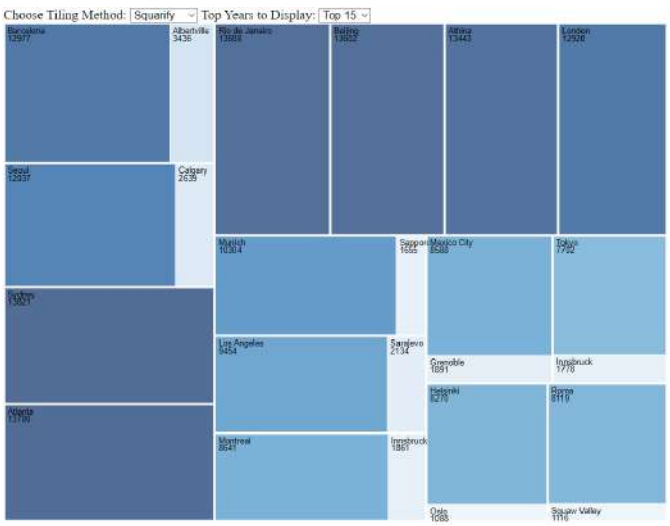


Fig. 14. Event Participation by Year and City Squarify method

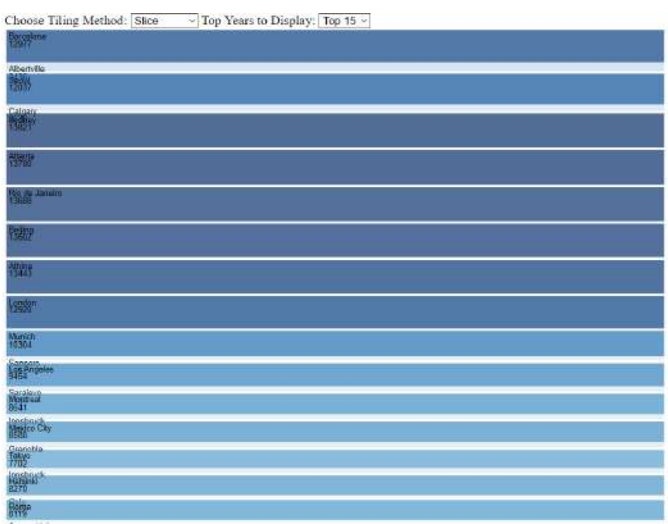


Fig. 15. Event Participation by Year and City Slice method

- **Implementation Summary:** After loading the data, the javascript code filters and groups entries by Year and City, then counts the number of

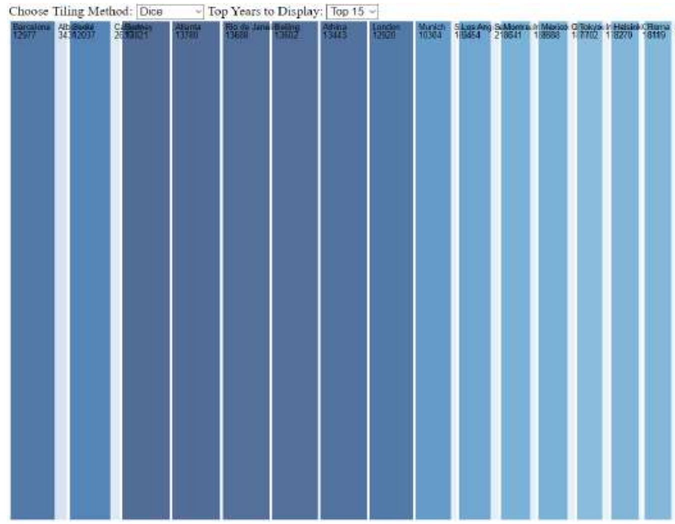


Fig. 16. Event Participation by Year and City Dice method

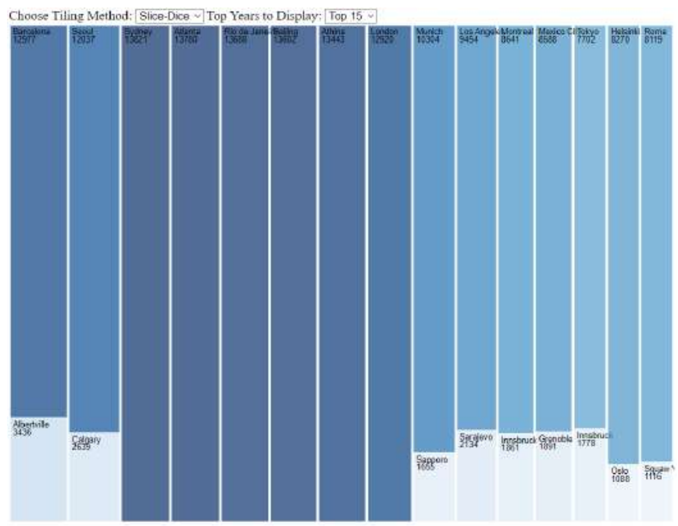


Fig. 17. Event Participation by Year and City Slice and Dice method

athletes in each group. The data is organized into a hierarchical structure compatible with the D3 treemap layout.
Users can select different tiling methods and display only the top N years with the most participants using dropdown menus.
There are user interactions for selecting :
  - Tiling method
  - No of Top Years(by Athlete count)

5) **Visualization 2: Medal Count by Country Across Sports:** This treemap visualization provides an insightful representation of Olympic medal counts by country across various sports. Each country is represented as a top-level block, subdivided by individual sports, where the size of each rectangle reflects the number of medals won. A color gradient from light to dark orange indicates higher medal counts,

making it easy to identify which countries dominate in particular sports.

Fig. 18. Medal Count by Country Across Sports Squarify method

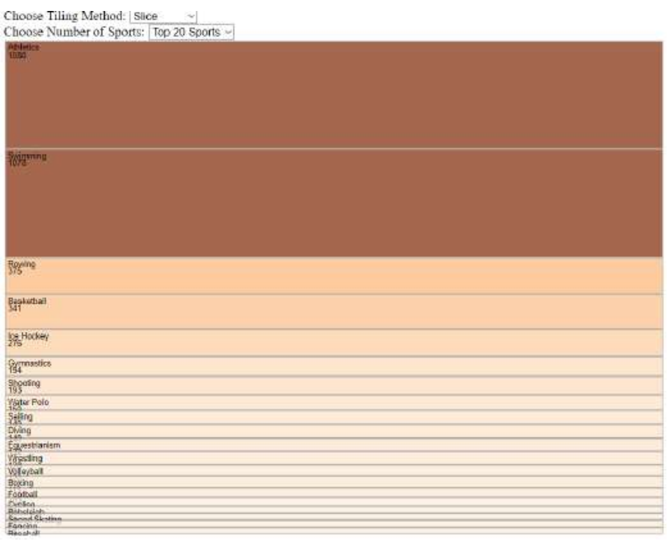


Fig. 19. Medal Count by Country Across Sports Slice method

- **Implementation Summary:** The code uses D3.js to create an interactive treemap that reads medal data from given dataset. It aggregates medal counts by country and sport, filtering out entries without medals. The data is structured hierarchically, allowing each country to contain multiple sports as children nodes. The visualization displays a **drop-down menu** for No of Sports too, making it manageable and focused.
  A dropdown menu enables users to switch between different treemap tiling methods ("Squarify," "Slice," "Dice," "Slice-Dice"), dynamically recalculating the layout.

Fig. 20. Medal Count by Country Across Sports Dice method

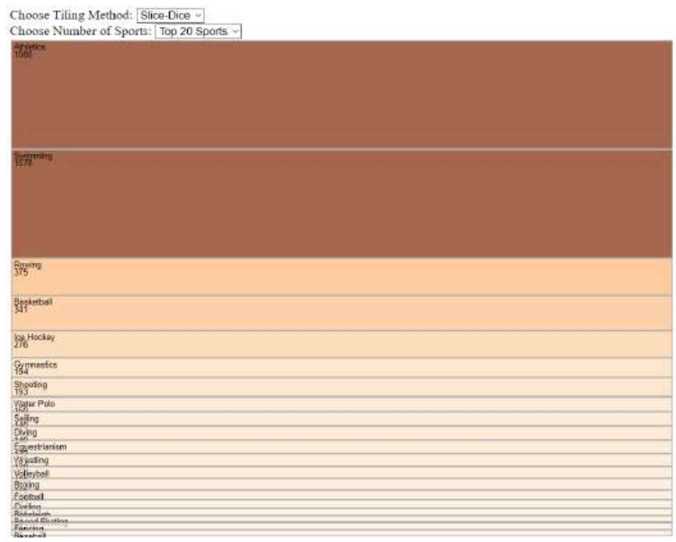


Fig. 21. Medal Count by Country Across Sports Slice and Dice method

6) **Visualization 3: Medal Distribution by Age Group and Sex** This creates an interactive treemap visualization to display the distribution of Olympic medals across different age groups and sexes. Using D3.js, the data is filtered and aggregated to show how many medals each age group within each sex category (male or female) has won.

   Key Insights:
   - The treemap can highlight differences in medal counts between males and females, shedding light on how certain age groups within each sex category outperform or underperform.
   - Clustering Patterns: The hierarchical nature of the treemap allows for the exploration of clustering patterns, showing whether certain combinations of

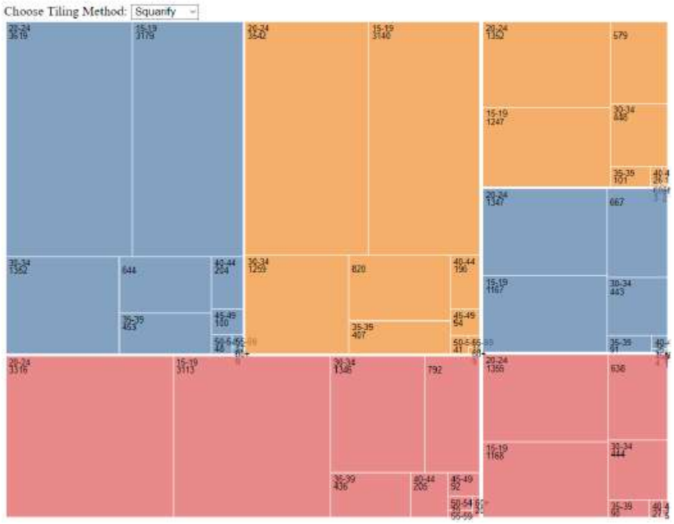


Fig. 22. Medal Distribution by Age Group and Sex Squarify method

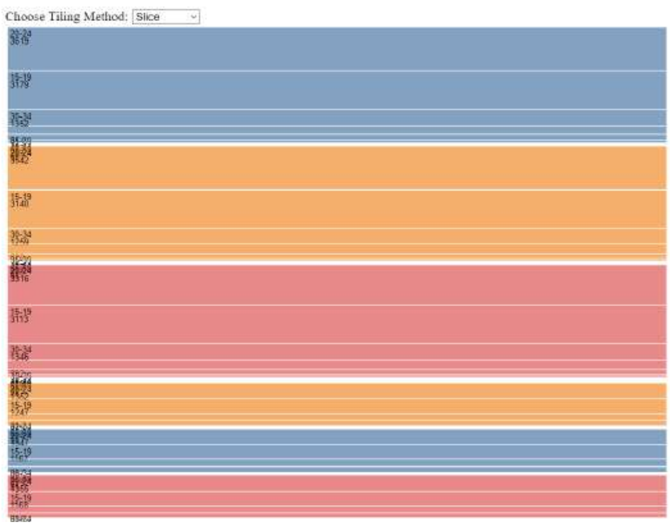


Fig. 23. Medal Distribution by Age Group and Sex Slice method

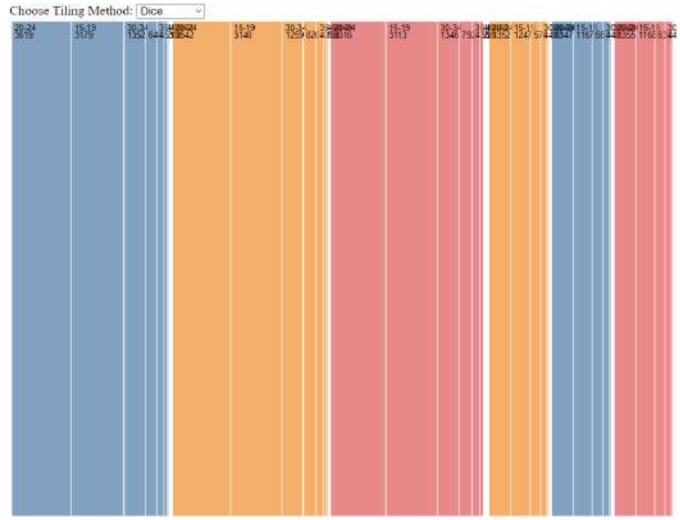


Fig. 24. Medal Distribution by Age Group and Sex Dice method

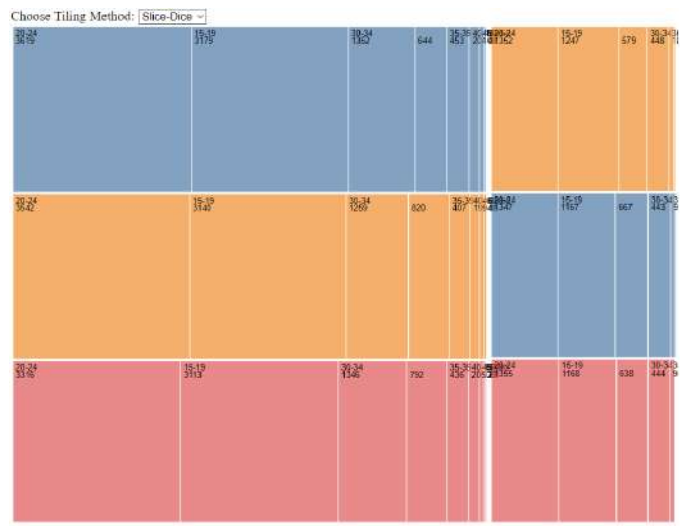


Fig. 25. Medal Distribution by Age Group and Sex Slice and Dice method

sex, age group, and medal type are more frequent than others.eg - Males in the age group of 20-24 have the most medals overall.

7) **Vizualization 4: Participation Over Time In a Particular Sports:** a treemap offers an insightful perspective on how the participation in that sport evolves over multiple Olympic Games. It efficiently displays hierarchical data and allows for the comparison of participation across different dimensions, such as time (years) and teams (countries or specific athletes).

III. Quiver Plots and Streamline Plots

A. Introduction

The wind analysis for the continental United States from November 2023 to January 2024 utilizes two datasets, `vs.nc` and `th.nc`, containing meteorological variables such as wind speed and wind direction. These datasets are processed in Python to generate quiver plots, visually representing wind direction and intensity over time and space.

The generated plots provide a comprehensive, biweekly overview of wind behavior over the study period, using two visualization styles:

- **Color-based glyphs**: In these plots, color variations represent wind speed, with arrows indicating direction.
- **Length-based glyphs**: In these plots, wind speed is represented by the length of arrows, while direction is conveyed by arrow orientation.

Each of these visualization techniques offers distinct insights, enabling a richer understanding of wind patterns across different regions and times. However, interpreting and comparing these quiver plots to draw inferences can be complex, requiring a detailed analysis of both temporal and spatial trends in wind characteristics.

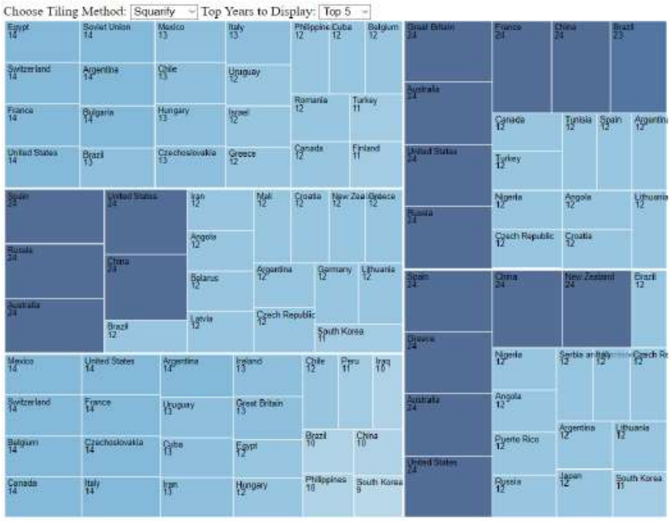


Fig. 26. Participation Over Time In a Particular Sports Squarify method

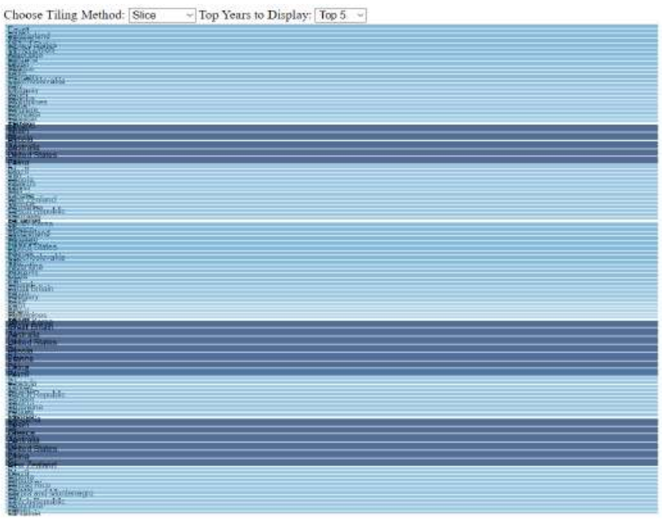


Fig. 27. Participation Over Time In a Particular Sports Slice method

Fig. 28. Participation Over Time In a Particular Sports Dice method

Fig. 29. Participation Over Time In a Particular Sports Slice and Dice method

B. Rationale Behind Choosing Time Period and Time Instances

The rationale for choosing the time period (November 2023 to January 2024) and the time instances (sampling every 15 days) for the quiver plot assignment is as follows:

- **Seasonal Variation:** The time period of November 2023 to January 2024 corresponds to the winter season in the northern hemisphere, specifically in the United States. During this period, wind patterns can show distinct seasonal behavior, such as changes in wind direction and speed due to atmospheric pressure shifts, which could provide valuable insights into wind dynamics during colder months.
- **Temporal Granularity:** By sampling every 15 days, a balance is struck between too frequent and too sparse data collection. This interval allows capturing key changes in wind behavior while avoiding excessive data points that may make the analysis cumbersome or result in too much noise. The 15-day interval also makes it easier to analyze potential bi-weekly or fortnightly trends and patterns.
- **Data Representativeness:** The dataset covering this period likely captures diverse weather phenomena, such as storm systems, pressure changes, and other atmospheric events, which affect wind direction and speed. The chosen time instances (every 15 days) help capture these variations effectively, ensuring the data is not too sparse to miss important events but also not overly frequent to lead to redundancy.
- **Geographic Relevance:** The US continent's wind patterns can vary across regions (coasts, plains, mountains), and focusing on a three-month period allows analyzing the wind's behavior in different areas with sufficient temporal coverage. Sampling every 15 days also provides

a good range of data points for geographical comparison without being too data-heavy for analysis.

- **Manageability:** With a 15-day sampling interval, the dataset remains manageable for analysis and visualization, especially when working with quiver plots. This approach provides enough data to observe trends over the winter period without overwhelming the visualization with too many points.

In summary, the chosen time period and sampling frequency enable an in-depth yet manageable analysis of wind behavior during the winter months, while considering both seasonal variations and practical constraints for data handling and visualization.

C. Objective

The objective of this analysis is to interpret the wind patterns across the U.S. from November 2023 to January 2024, identify key seasonal trends, compare the efficacy of color-based and length-based quiver plot visualizations, and derive meaningful meteorological insights from the observed data. This study also aims to address the challenges and complexities inherent in interpreting and comparing multiple visualization styles over an extended timeframe.

D. Actions

The following steps were undertaken to generate the wind plots and conduct the analysis:

1) Data from the `vs.nc` and `th.nc` files was loaded and processed using Python libraries.
2) Wind speed and direction were visualized using color-based and length-based quiver plots.
3) Biweekly snapshots were generated to capture temporal variations across the November to January period.
4) A comparative analysis was conducted between the two visualization styles to assess their respective strengths and limitations in illustrating wind patterns.

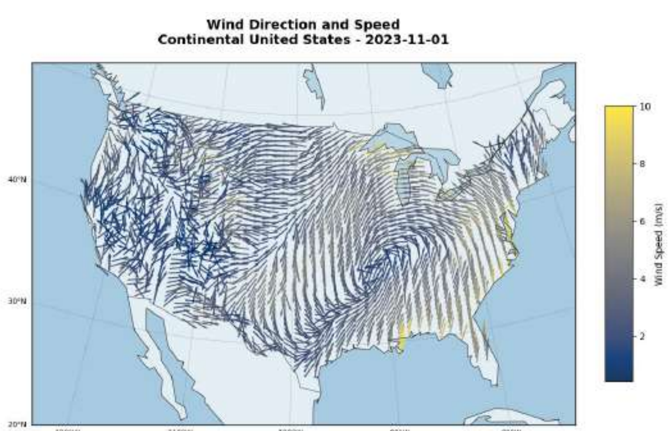


Fig. 30. Color-Based Quiver Plot (2023-11-01): Wind speed by color intensity; arrows indicate direction.

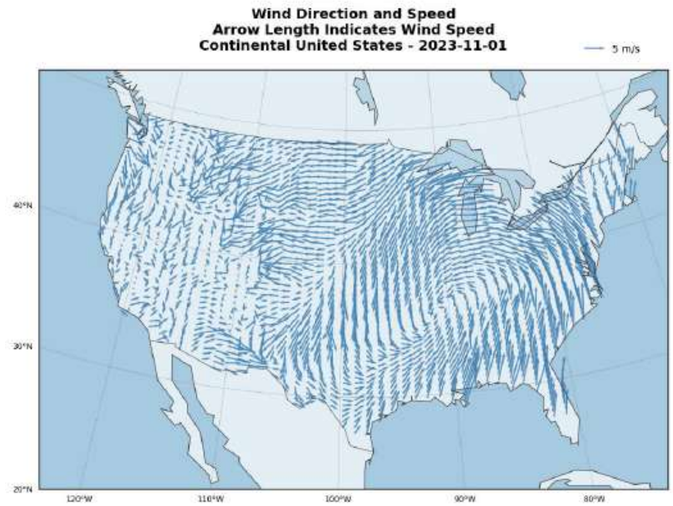


Fig. 31. Length-Based Quiver Plot (2023-11-01): Wind speed by arrow length; direction shown by orientation.

E. Methodology and Preprocessing

Data Sources: The data set comprises NetCDF files containing daily wind speed and direction values for the specified period. Separate files for wind direction and wind speed are utilized for each year.

Preprocessing Steps:

- Extract latitude, longitude, and wind parameters from NetCDF files.
- Convert wind direction values from degrees to radians to compute U and V vector components.
- Sample the dataset to limit the number of arrows displayed, enhancing plot clarity.
- Normalize vectors for uniform quiver plot display when using color as the encoding parameter for wind speed.

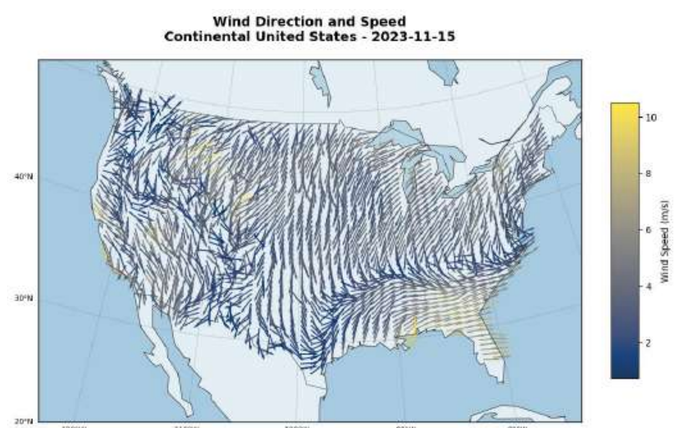


Fig. 32. Color-Based Quiver Plot (2023-11-15): Wind speed by color intensity; arrows indicate direction.

F. Implementation

To implement the visualizations:

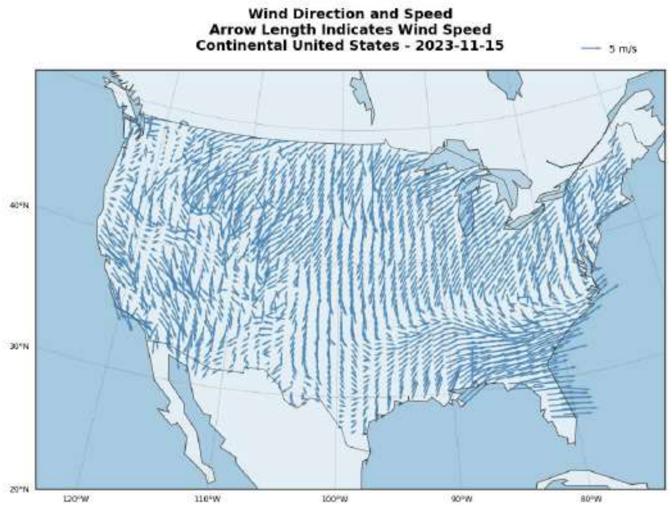


Fig. 33. Length-Based Quiver Plot (2023-11-15): Wind speed by arrow length; direction shown by orientation.

- Define functions for both quiver plots and streamline plots.
- For each selected date, generate static images with wind direction and speed visualized through quiver arrows. For color-based visualizations, color reflects wind speed, while length remains constant. For length-based visualizations, the length of arrows is scaled based on wind speed.
- Save each plot as an image file, which is later compiled into a GIF to observe wind pattern changes over time.

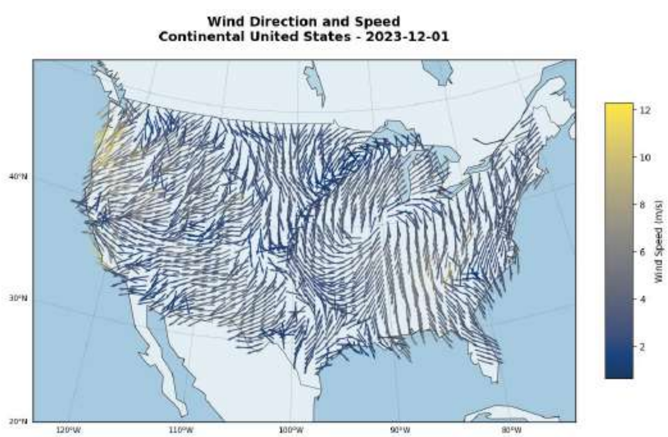


Fig. 34. Color-Based Quiver Plot (2023-12-01): Wind speed by color intensity; arrows indicate direction.

G. Detailed Inferences from the Plots

1) General Wind Pattern Observations:

- **Western United States**: The color-based plots reveal stronger winds in the northwestern coastal areas, with speed generally decreasing toward the interior. The length-based plots further clarify this pattern, showing shorter arrows inland, indicating lower wind speeds.

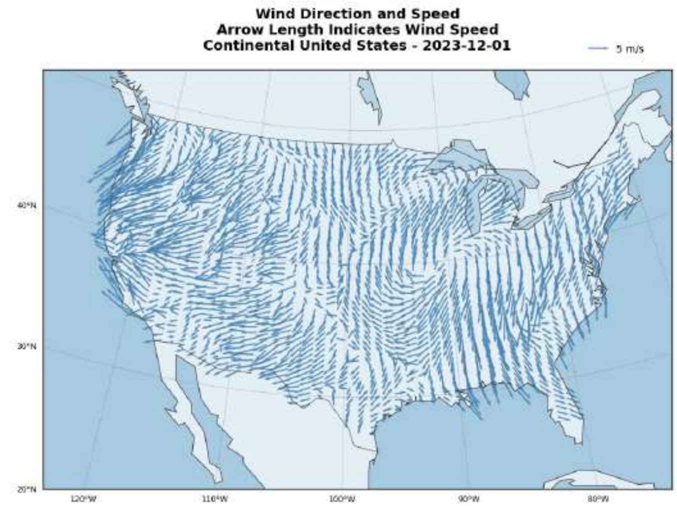


Fig. 35. Length-Based Quiver Plot (2023-12-01): Wind speed by arrow length; direction shown by orientation.

- **Central Plains**: Consistent high wind speeds are observed in the central plains, especially in the northern Great Plains region. Both visualization methods show convergence patterns here, indicating a zone where different air masses meet, which is typical of winter months when polar air descends southward.
- **Eastern Coast**: The eastern seaboard often shows higher wind speeds, particularly in the mid-Atlantic and Northeast regions, influenced by winter storms moving along the coast. The color-based plots capture rapid changes in speed well, showing streaks of yellow in high-speed zones, while the length-based plots emphasize the directional consistency.

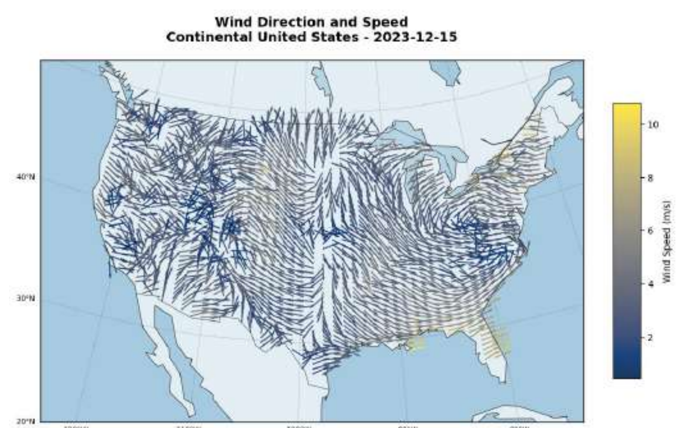


Fig. 36. Color-Based Quiver Plot (2023-12-15): Wind speed by color intensity; arrows indicate direction.

2) Temporal Changes and Seasonal Trends:

- **November to December Transition**: The wind speeds in November (color-based) show lower speeds across the southern U.S., with cooler northern areas showing moderate activity. By mid-December, there is an increase in

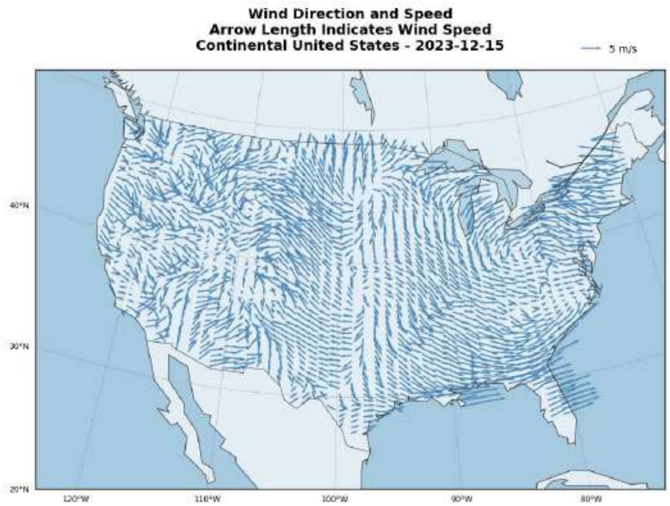


Fig. 37. Length-Based Quiver Plot (2023-12-15): Wind speed by arrow length; direction shown by orientation.

wind speeds across the central and northeastern regions, indicative of stronger winter storms.

- **December to January Transition**: In the December–January period, both plot types indicate a marked increase in wind intensity across most of the U.S., especially along the east coast and Midwest, possibly due to intensified winter storms. The length-based plots make it easy to observe changes in average wind speeds, with longer arrows becoming more prevalent across wider areas.

Fig. 38. Color-Based Quiver Plot (2024-01-01): Wind speed by color intensity; arrows indicate direction.

3) Comparing Visualization Styles:

- **Color-Based Glyphs**: These plots excel at highlighting regions with extreme wind speeds due to the clear color contrast (e.g., blue for lower speeds, yellow for higher speeds). They are particularly useful for identifying isolated areas with very high or low speeds and can show how wind speeds vary within a single region.

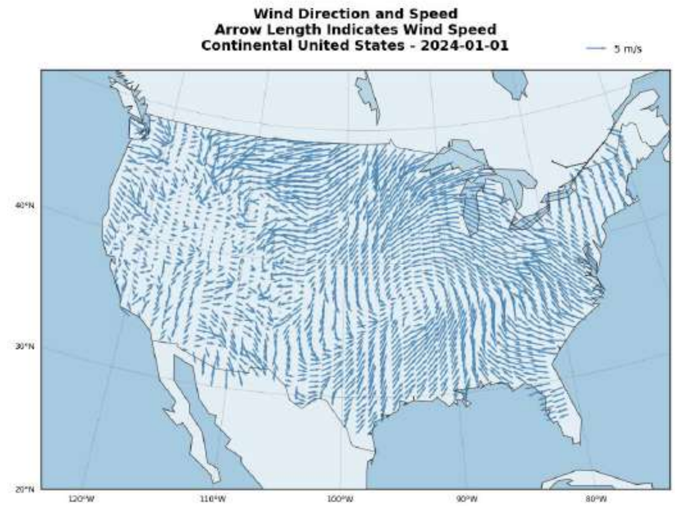


Fig. 39. Length-Based Quiver Plot (2024-01-01): Wind speed by arrow length; direction shown by orientation.

- **Length-Based Glyphs**: These plots provide an intuitive sense of speed and direction across broad areas. They emphasize overall wind patterns more than localized extremes, as the length variation is subtler than color shifts. This style is beneficial for quickly understanding directional trends and is particularly effective in identifying regions with consistent wind speeds.

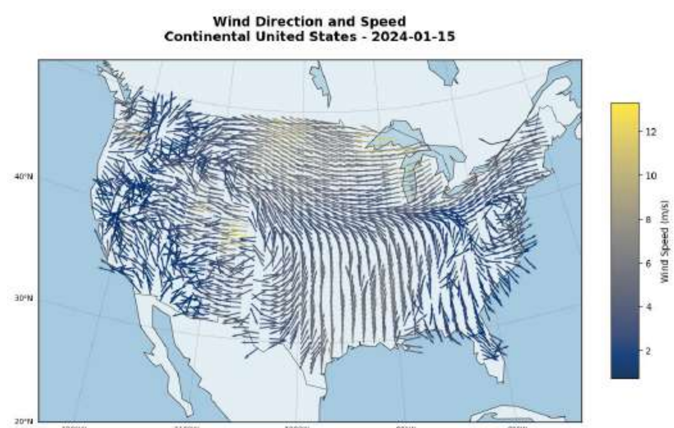


Fig. 40. Color-Based Quiver Plot (2024-01-15): Wind speed by color intensity; arrows indicate direction.

4) Regional Focus and Seasonal Effects:

- **Mountainous and Coastal Regions**: Both visualization types show significant wind variability in the Rocky Mountain region, where complex topography affects wind patterns. Coastal regions, especially in the Pacific Northwest and Northeast, exhibit consistently strong winds throughout the period, likely influenced by oceanic airflows and winter storm paths.
- **Midwestern States**: A recurrent cyclonic pattern is visible in the color-based plots, particularly in December and January, over the Midwest. This pattern suggests a

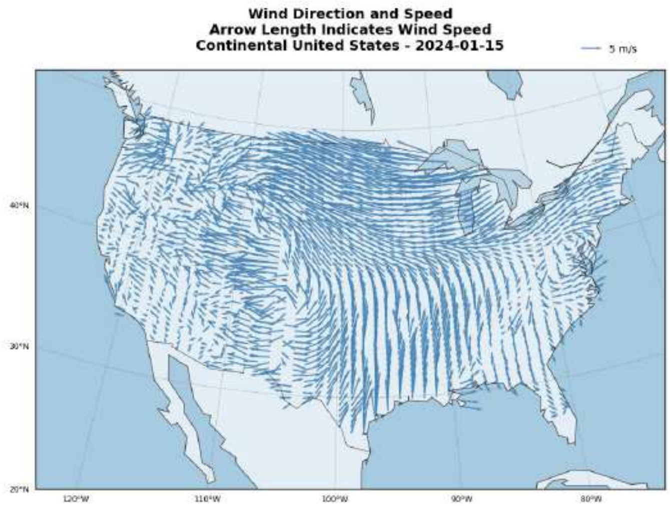


Fig. 41. Length-Based Quiver Plot (2024-01-15): Wind speed by arrow length; direction shown by orientation.

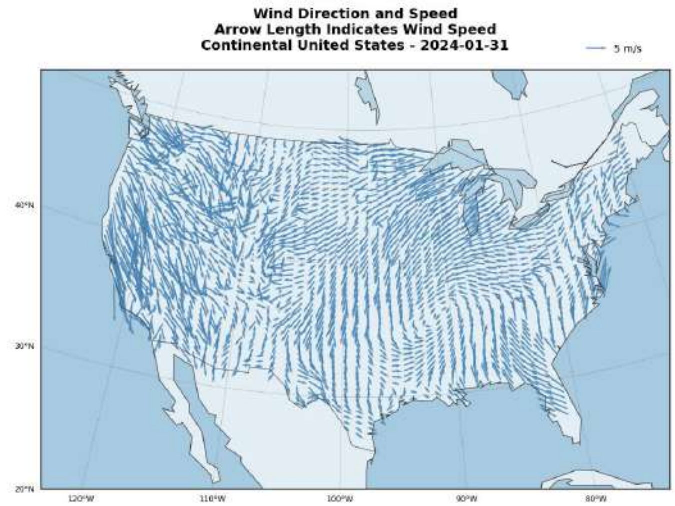


Fig. 43. Length-Based Quiver Plot (2024-01-31): Wind speed by arrow length; direction shown by orientation.

stationary or slow-moving low-pressure system typical of winter months, where cold and warm fronts frequently interact.

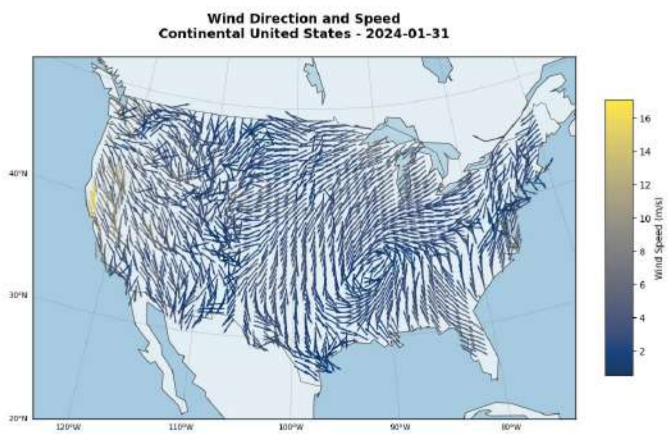


Fig. 42. Color-Based Quiver Plot (2024-01-31): Wind speed by color intensity; arrows indicate direction.

H. Complexity in Arriving at Inferences

- **Data Interpretation Challenges**: Both color and length-based plots require careful interpretation. Color gradients can sometimes obscure details if not viewed in high resolution, and length variations may be difficult to discern in densely packed arrow regions. Comparing these different glyph styles adds another layer of complexity, as each style emphasizes different aspects of wind behavior.
- **Temporal Analysis**: Observing trends over multiple intervals (14 plots, spanning three months) involves distinguishing between random variations and actual seasonal patterns. Weather systems vary on daily and weekly timescales, so inferring meaningful trends from biweekly snapshots demands careful scrutiny, often requiring additional meteorological knowledge.
- **Regional Variation in Wind Behavior**: Due to the complex terrain and diverse climates in the U.S., wind patterns vary significantly by region. Understanding the reasons behind these variations, such as coastal influence, topography, or interaction of polar and tropical air masses, is essential for accurate inferences. This requires familiarity with geographical and atmospheric science concepts.
- **Technical Complexity**: Processing the `vs.nc` and `th.nc` files to create these quiver plots in Python involves specialized knowledge of netCDF data structures and plotting libraries (e.g., Matplotlib, Cartopy). Each plot must be carefully configured to ensure accurate representation of wind speed and direction.
- **Visualization Limitations**: Both glyph types have their limitations. The color-based plots, while effective for distinguishing high-speed areas, can sometimes make it difficult to gauge exact direction if colors are very close. Length-based glyphs can give a clearer sense of directional flow, but subtle speed changes may be lost, especially in regions with similar wind speeds.

1) Complexity in Overlapping Regions:

- **Length-based plot on December 15**: The Great Plains and Northeast are particularly challenging regions due to vector overlap. High-density areas make it difficult to observe small directional shifts as vectors overlap.
- **Color-based plot on January 15**: High-density areas are more interpretable using color gradients. The color-based glyphs allow for easy differentiation of speeds even in complex regions where vectors overlap or wind direction shifts frequently.

2) Seasonal Patterns and Changes Over Time:

- **Length-based sequence (November to January)**: Ob-

serving vector length changes across dates shows a general increase in wind speed in the northern regions, especially near the Great Lakes and New England, reflecting typical winter storm patterns.
- **Color-based sequence (November to January)**: Color intensity changes over time allow for an effective visualization of speed fluctuations. In late January, intense colors in the Midwest highlight high-speed areas that align with late-winter storms.

I. Wind Patterns and Observations

1) Early Winter Patterns (November 1 and 15): In early winter, strong wind patterns begin to form across the Midwest and Great Plains, influenced by developing storm systems.

- **Length-based plot on November 1**: Long vectors appear prominently in the Midwest and Great Plains, indicating strong winds typical of early winter storms. The Pacific Northwest and Northern Plains also show extended vectors, likely due to cold air masses moving through these regions.
- **Color-based plot on November 1**: The color-based plot highlights high-speed zones through warm colors, particularly in northern regions and along the eastern seaboard. These colors reflect potential oceanic influences on wind speed near the coast, which is more easily seen through the color gradient.

2) Mid-Winter Trends (December 1 and 15): By mid-winter, wind speeds typically intensify across northern regions as colder air masses push southward.

- **Length-based plot on December 1**: Vector lengths around the Great Lakes region indicate a significant increase in wind strength, likely due to cold fronts moving southward. The Rockies also show long vectors, reflecting strong winds as cold air masses encounter mountainous terrain.
- **Color-based plot on December 1**: The color gradient highlights wind intensity across both coastal and central regions, with warm colors indicating high speeds. This makes wind activity in urban areas, such as the Northeast, easy to interpret despite overlapping vectors.

3) Late Winter Trends (January 1, 15, and 31): Late winter brings a final increase in wind intensity as strong fronts and jet streams impact the Midwest and East Coast.

- **Length-based plot on January 15**: There is a clear pattern of sustained high-speed winds along the East Coast, indicated by longer vectors. This pattern aligns with late-winter cold fronts, which tend to bring powerful winds. Conversely, shorter vectors appear in the Southeast, suggesting calmer conditions in this area.
- **Color-based plot on January 31**: The color gradient indicates a high-speed corridor extending from the Midwest down to the Southern Plains. The warm color intensity in this corridor reflects the typical strengthening of the jet stream in late winter, with enhanced visibility for smaller, turbulent areas.

J. Comparative Analysis of Length-Based and Color-Based Visualizations

Both visualization methods offer distinct advantages and limitations. Below is a comprehensive comparison:

- **Speed Representation**:
  - *Length-Based Quiver Plots*: Wind speed is represented by vector length, with longer vectors indicating higher speeds. This approach is effective for quickly identifying high-wind areas in open regions.
  - *Color-Based Quiver Plots*: Wind speed is represented by color intensity, with warmer colors signifying higher speeds. This method provides clear visual feedback, especially in densely populated or high-overlap regions.
- **Directional Clarity**:
  - *Length-Based Quiver Plots*: In regions with high vector density, overlapping vectors can obscure directional clarity, particularly where wind patterns shift frequently. This issue is most prominent in the Northeast and Midwest regions.
  - *Color-Based Quiver Plots*: The consistent vector length improves directional clarity, even in areas where vectors overlap or where high vector density might otherwise obscure wind direction.
- **Regional Utility**:
  - *Length-Based Quiver Plots*: This visualization works best in open regions with low vector density, such as the Great Plains and Rocky Mountains, where vector overlap is minimal and length variation is easy to interpret.
  - *Color-Based Quiver Plots*: The color gradient is advantageous in densely populated or high-overlap areas, like the Northeast and Midwest, where color-based speed representation improves visibility and interpretation.
- **Temporal Analysis**:
  - *Length-Based Quiver Plots*: Seasonal wind changes are noticeable through variations in vector length over time, effectively indicating speed increases or decreases, especially in the northern and central regions.
  - *Color-Based Quiver Plots*: Gradual shifts in color over time provide an effective way to track seasonal changes in speed, particularly in areas with complex or high-density wind patterns.

From this comparison, it's evident that while length-based plots offer a straightforward representation of wind speed, color-based plots provide greater clarity in high-density areas, allowing for easier speed differentiation without sacrificing directional information.

K. Introduction to Streamline Plots

Streamline plots visualize flow patterns in a fluid field by representing the direction and strength of flow. They show the path that a fluid element would follow at each point, with denser lines indicating regions of stronger flow. These plots are particularly useful in fluid dynamics, meteorology, and environmental modeling to analyze air, water, or gas movement across regions or over time.

L. Detailed Inferences and Observations

Each streamline plot reflects the flow characteristics on different days, demonstrating changes in intensity, direction, and structural patterns. By examining each day's plot, we can infer insights about the flow field's dynamic behavior, helping us understand any transient systems, potential pressure changes, or areas of instability.

- **Day 09** (Figure 44): Observed stable flow with smooth, widely spaced streamlines, indicating low velocity and low-pressure gradient in the flow field. This could correspond to a region experiencing calm, stable conditions without significant disturbances.

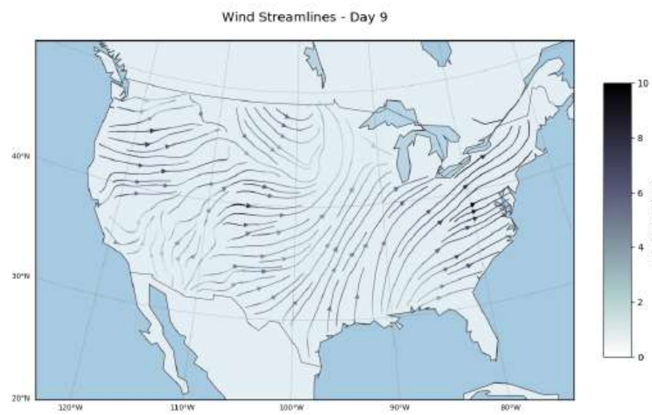


Fig. 44. Streamline Plot for Day 09

- **Day 19** (Figure 45): Slight curvature intensification and some increase in streamline density are seen, which may signify a building flow system. This suggests a moderate increase in flow velocity or an impending weather shift in the context of atmospheric modeling.
- **Day 29** (Figure 46): Convergence of streamlines shows denser and more curved patterns. This indicates an area of low pressure or a developing disturbance, suggesting that the system is evolving toward a higher energy state, potentially signaling a storm or atmospheric front.
- **Day 304** (Figure 47): High-density streamlines with turbulent patterns indicate a high-energy flow, typical of significant weather events. This day shows stronger interactions between flow lines, suggesting a high-pressure difference driving rapid changes in velocity and direction.
- **Day 314** (Figure 48): The flow begins to stabilize, with streamlines spreading slightly apart. This could imply that

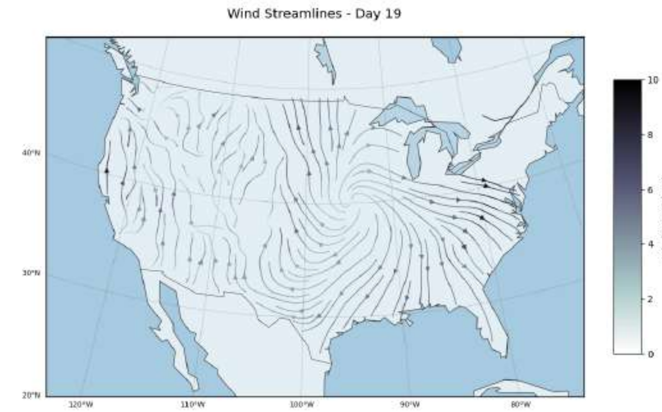


Fig. 45. Streamline Plot for Day 19

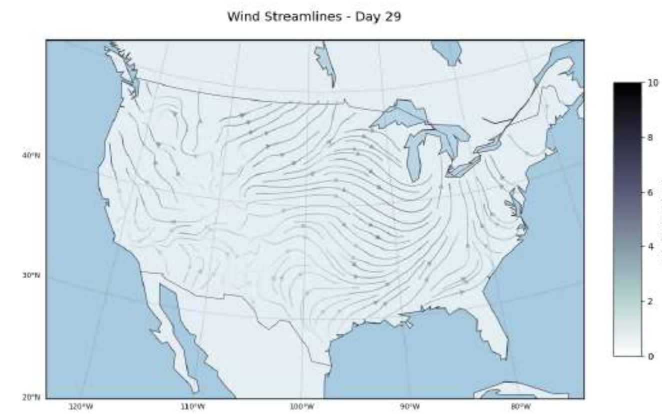


Fig. 46. Streamline Plot for Day 29

the peak intensity of the previous day's system is reducing, as flow velocity decreases and pressure gradients lessen.
- **Day 324** (Figure 49): A change in streamline orientation shows a marked directional shift, potentially indicating a change in the primary flow direction. This type of shift often precedes larger atmospheric or environmental changes.
- **Day 334** (Figure 50): The flow re-stabilizes, and the streamline pattern becomes smoother and more parallel. This indicates a return to steadier conditions, likely representing a high-pressure, low-turbulence system.
- **Day 344** (Figure 51): The plot reveals some minor disturbances, with localized intensifications within an otherwise stable field. These fluctuations may signify small-scale weather events or localized flow disturbances that impact

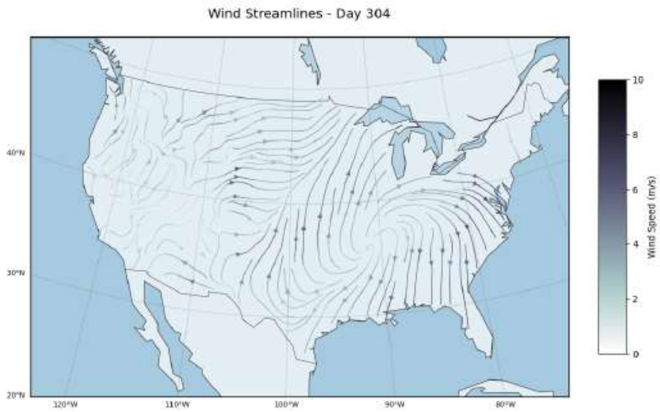


Fig. 47. Streamline Plot for Day 304

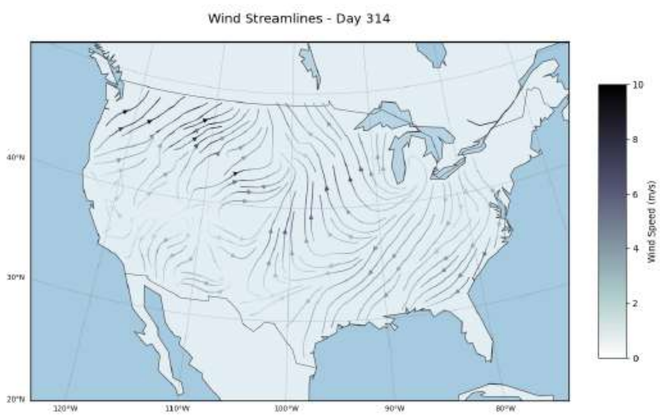


Fig. 48. Streamline Plot for Day 314

the main system only minimally.

- **Day 354** (Figure 52): Streamlines begin to gather and curve again, suggesting the early stages of another high-energy event or atmospheric front. This could be an indicator of a developing system or a new flow instability in the environment.
- **Day 364** (Figure 64): The flow appears smooth and widely spaced once more, reflecting stable, low-turbulence conditions. This day likely represents a calm period, possibly after a series of disturbances.

M. Complexity of Arriving at Inferences

Analyzing streamline plots requires an understanding of fluid dynamics, as each flow line and its behavior reflect specific environmental conditions:

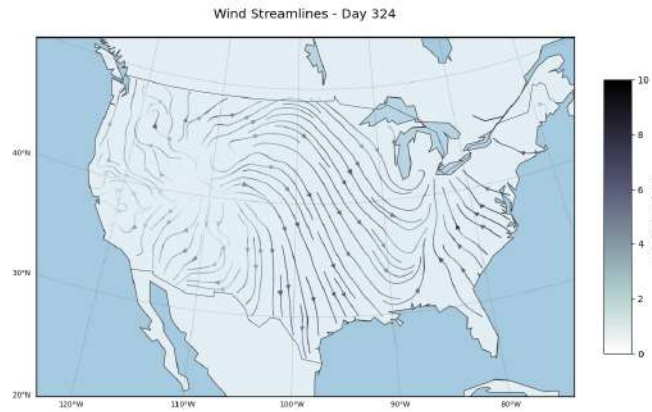


Fig. 49. Streamline Plot for Day 324

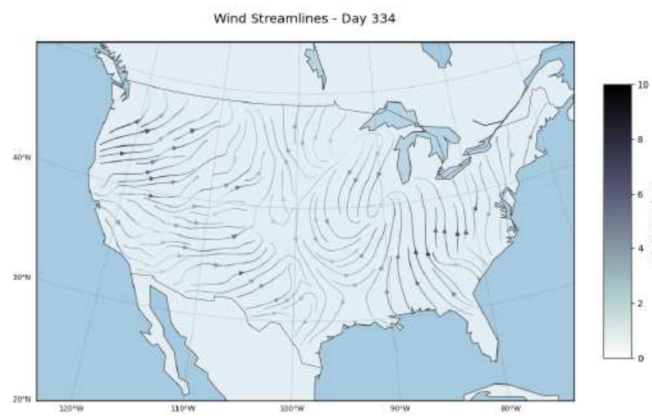


Fig. 50. Streamline Plot for Day 334

- **Interpreting Curvatures and Density**: Increased streamline density indicates faster flow regions or areas with higher velocity gradients, often pointing toward turbulence. Curvatures imply changes in flow direction due to external forces, which are not always immediately clear without contextual data.
- **System Evolution and Pattern Recognition**: Understanding how systems evolve across sequential days requires observing subtle shifts in density and direction. A trained eye is necessary to interpret these shifts, as it's not just a matter of identifying high or low flow regions, but also understanding how these regions interact over time.
- **Understanding Flow Stability**: Judging stability versus turbulence from streamline plots alone requires the ability to distinguish between naturally fluctuating flows and

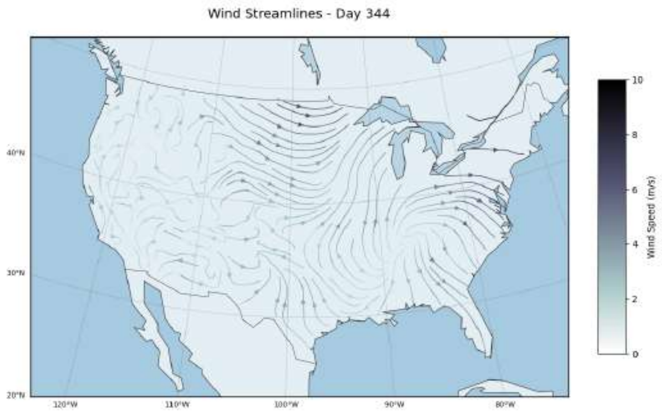


Fig. 51. Streamline Plot for Day 344

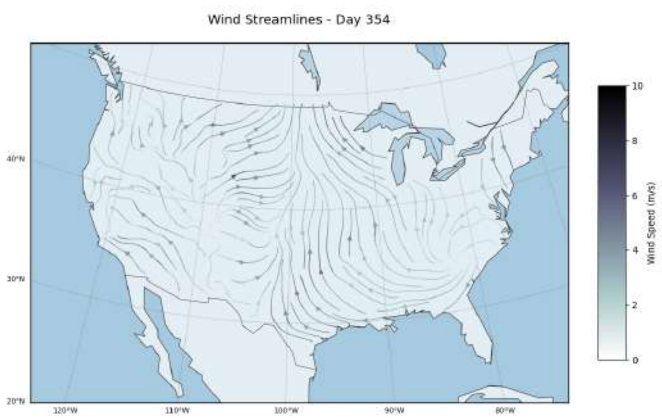


Fig. 52. Streamline Plot for Day 354

actual high-energy disturbances. This requires familiarity with typical patterns in the specific field (e.g., meteorology, oceanography).

N. Comparison with Quiver Plots

Unlike streamline plots, quiver plots represent flow using arrows, where each arrow shows the direction and magnitude of flow at a particular point.

- **Advantages of Streamline Plots**: Streamline plots provide a continuous visualization of flow, making it easier to follow flow paths and identify coherent structures within the flow, such as vertices or areas of convergence.
- **Advantages of Quiver Plots**: Quiver plots can more accurately depict magnitude changes at specific points, which is beneficial for localized analysis. However, quiver

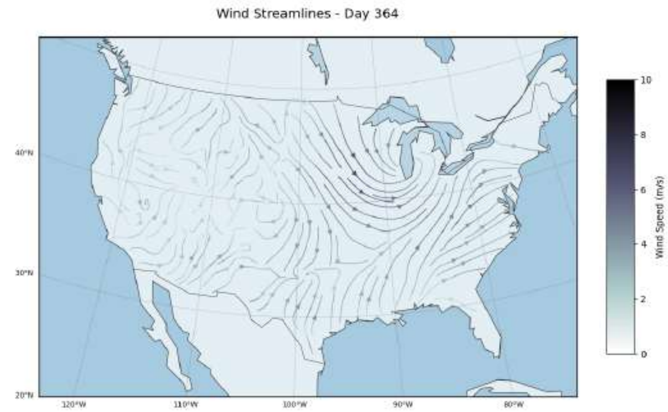


Fig. 53. Streamline Plot for Day 364

plots can become cluttered in high-density areas, making it difficult to visualize overall patterns.
- **Use Case Differences**: Streamline plots are often preferred for large-scale flow visualizations where smoothness and path continuity are important, whereas quiver plots are ideal for detailed, localized studies.

O. Conclusion

The sequence of streamline plots provided a detailed view of the fluid flow changes over the period. Observing the shifts in flow intensity and direction allows for identifying transient weather events or systemic changes in the environment. For instance, days like 304 and 354 indicate higher turbulence or stronger flow systems, while days like 334 and 364 represent stable conditions. These analyses contribute significantly to forecasting models and environmental assessments.

P. Conclusions

- **Seasonal Shift in Wind Patterns**: Both glyph styles reveal seasonal shifts well, with length-based plots highlighting speed changes in early winter (November-December) and color-based plots emphasizing variations in late winter (January).
- **Regional Wind Dynamics**: Both glyphs capture heightened wind activity from November through January, with increased wind speeds in the northern and central US regions due to winter storms.
- **Glyph Strengths by Region**: Length-based glyphs are particularly suited for open regions with minimal overlap, such as the Great Plains, whereas color-based glyphs provide enhanced clarity in high-density regions, such as the Northeast and Midwest.

IV. Parallel Coordinates Plot Analysis

A. Introduction

Parallel Coordinates Plots (PCP) are effective tools for visualizing and exploring complex, multi-dimensional datasets. They are particularly useful in detecting patterns, correlations, and anomalies across a large number of variables. In this report, we analyze two interactive PCPs:

- **Jazz Network PCP**: This plot represents collaboration patterns among jazz musicians, capturing the strength and distribution of connections within a musical network.
- **Olympic Athletes PCP**: This plot visualizes attributes of Olympic medalists, allowing exploration of physical and demographic trends across athletes from different sports.

Interactive features like brushing and axis reordering in these PCPs enhance their analytical capabilities, allowing users to engage deeply with the data. Through these features, we gain insights that would be challenging to obtain through traditional charts.

B. Jazz Network Parallel Coordinates Plot

1) Dataset Description: The Jazz Network dataset consists of data about musicians and their collaboration intensities. The key variables in this dataset are:

- **Node1** and **Node2**: Represent individual musicians, where a connection between nodes indicates a collaborative performance or recording.
- **Weight**: Reflects the strength or frequency of collaborations between musicians, with higher values indicating more frequent or intense collaborations.

2) Interactive Features: Brushing and Axis Reordering: The Jazz Network PCP incorporates interactive features that facilitate focused data exploration:

- **Brushing**: Brushing enables users to highlight ranges of values on specific axes, such as isolating high **Weight** values to focus on strong collaborations. By brushing along different axes, users can explore collaboration patterns for various groups or individuals within the jazz community.
- **Axis Reordering**: This feature allows users to rearrange axes, helping to reveal direct correlations between variables. For instance, by placing the **Weight** axis next to **Node1** or **Node2**, users can better understand how collaboration strength relates to specific musicians.

3) Objective and Significance: The Jazz Network PCP provides a visual representation of social and professional networks within jazz. The analysis objectives include:

- Identifying clusters of musicians with strong collaboration ties, suggesting influential groups or stylistic affiliations.
- Pinpointing central musicians who bridge multiple groups, highlighting key figures who may influence the network.
- Analyzing variations in connection strength, potentially revealing patterns of occasional versus frequent collaborations.

The ability to interactively filter and reorder variables enhances the user's ability to analyze network dynamics, making it possible to derive insights into the structure and central figures within the jazz community.

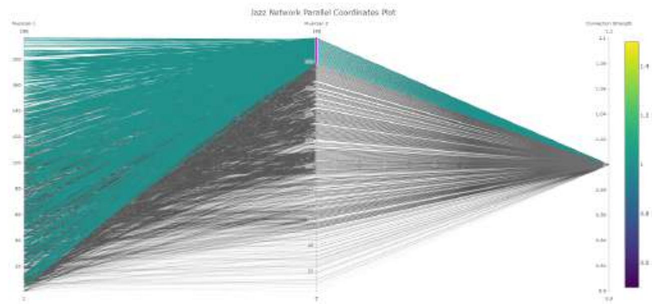


Fig. 54. Parallel Coordinates Plot for Jazz Networks Dataset after brushing

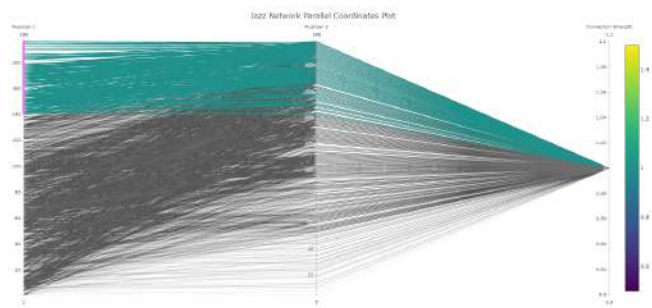


Fig. 55. Parallel Coordinates Plot for Jazz Networks Dataset after brushing

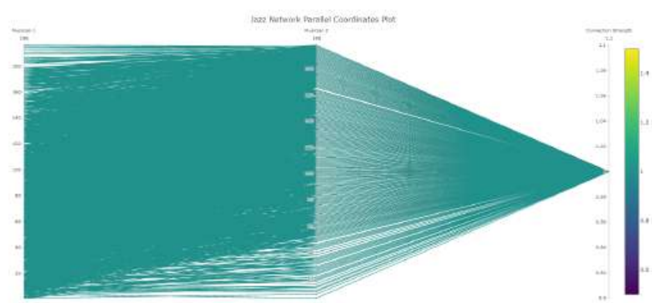


Fig. 56. Parallel Coordinates Plot for Jazz Networks Dataset

4) Inferences and Observations: The Jazz Network PCP allows us to infer several important insights:

- **Collaboration Clusters**: By brushing to isolate high-weight connections, dense clusters of collaborations are revealed. These clusters likely represent groups of musicians who frequently collaborate, possibly due to shared musical styles or frequent joint performances.
- **Central Figures in the Network**: Axis reordering helps identify musicians who serve as central nodes, acting

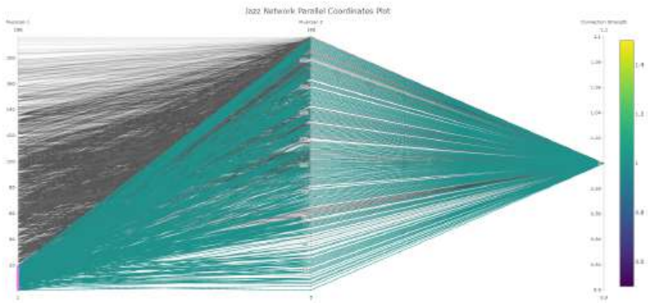


Fig. 57. Parallel Coordinates Plot for Jazz Networks Dataset after brushing

as bridges between various groups. These musicians are likely to be highly influential, contributing to stylistic developments or bridging different jazz subgenres.

- **Variable Collaboration Strengths**: We observe varying degrees of collaboration strength. High-weight connections indicate long-term partnerships, while lower weights may suggest one-off or less frequent collaborations. This differentiation provides insights into the nature and depth of professional relationships in the jazz community.
- **Subgroup Dynamics**: Brushing across different collaboration strength ranges also helps explore smaller, less prominent subgroups, which may represent unique stylistic subcultures within jazz.

5) Complexities and Disadvantages: While the Jazz Network PCP provides valuable insights, it also presents some challenges:

- **Overlapping Lines**: With a large number of collaborations visualized simultaneously, lines representing connections can overlap, making it difficult to distinguish individual relationships, especially when there are many low-weight connections.
- **Interpretation of Categorical Nodes**: Since musicians are represented by categorical identifiers, the numerical order of nodes does not convey intrinsic meaning. Understanding the network requires additional information about each musician, which is not provided in the PCP alone.
- **Complexity of Cluster Analysis**: Identifying clusters requires careful brushing and reordering, which can be time-consuming and challenging to interpret if clusters are not well-defined or if they have loose connections.

6) Conclusion: The Jazz Network PCP, with interactive brushing and axis reordering, offers a rich exploration of collaboration patterns and social structures within jazz. By allowing users to filter data and reorder axes, the PCP enables detailed examination of collaboration intensities, central figures, and network clusters. However, the high density of connections and categorical nature of musician identifiers present interpretative challenges, particularly in distinguishing subtle connections and interpreting musician identity from node data.

C. Olympic Athletes Parallel Coordinates Plot

1) Dataset Description: The Olympic Athletes dataset captures demographic and physical attributes of athletes who have won Olympic medals. Key variables include:

- **Year**: Year of the Olympic Games.
- **Age**, **Height**, and **Weight**: Physical characteristics of athletes.
- **Gender (0=M, 1=F)**: Encodes gender, allowing analysis across male and female athletes.
- **Team** and **Sport**: Categorical codes representing the athlete's team and sport.
- **Medal Category**: Indicates medal type (0 = Bronze, 1 = Silver, 2 = Gold).

2) Interactive Features: Brushing and Axis Reordering: The Olympic Athletes PCP uses brushing and axis reordering for detailed examination:

- **Brushing**: Users can highlight specific ranges along any axis, allowing exploration of athletes within certain age, height, or weight brackets. For instance, brushing along **Height** can reveal height distributions by sport, while brushing along **Medal Category** helps analyze characteristics associated with specific medal levels.
- **Axis Reordering**: Rearranging axes allows users to explore direct relationships, such as between **Sport** and **Medal Category** or **Gender** and **Weight**. This flexibility enables users to identify demographic and physical trends associated with medal success.

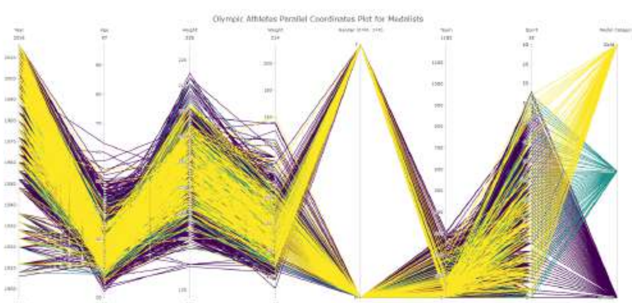


Fig. 58. Parallel Coordinates Plot for Olympic Dataset

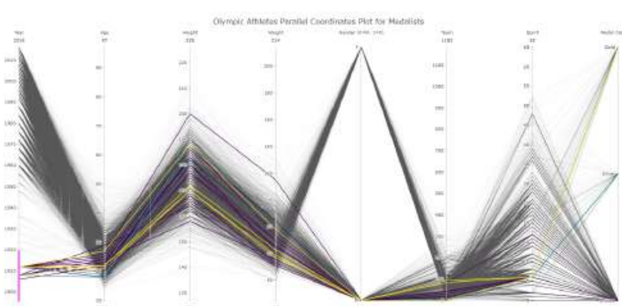


Fig. 59. Parallel Coordinates Plot for Olympic Dataset after brushing

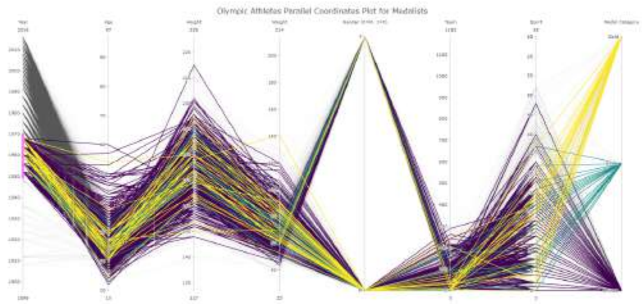


Fig. 60. Parallel Coordinates Plot for Olympic Dataset after brushing

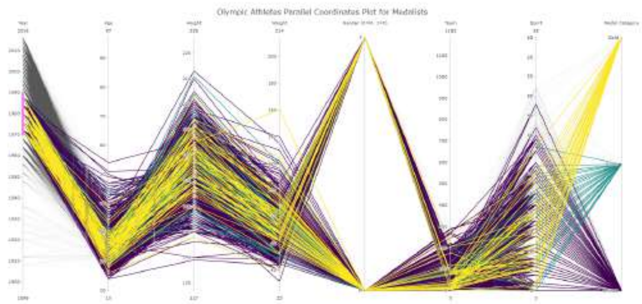


Fig. 61. Parallel Coordinates Plot for Olympic Dataset after brushing

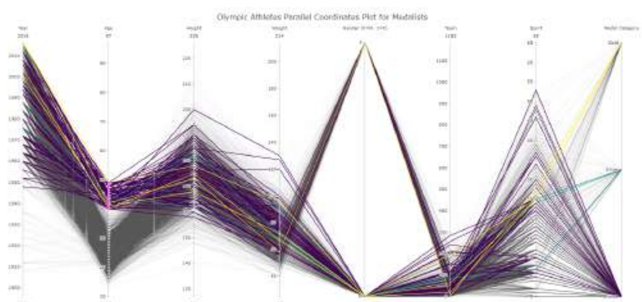


Fig. 62. Parallel Coordinates Plot for Olympic Dataset after brushing

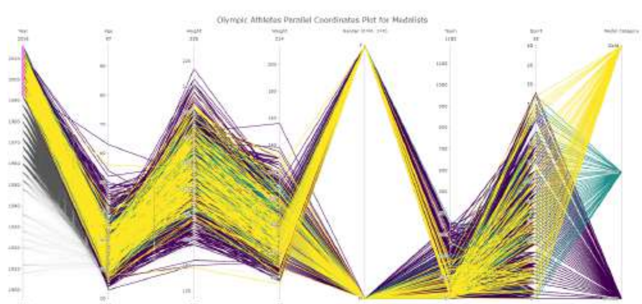


Fig. 63. Parallel Coordinates Plot for Olympic Dataset after brushing

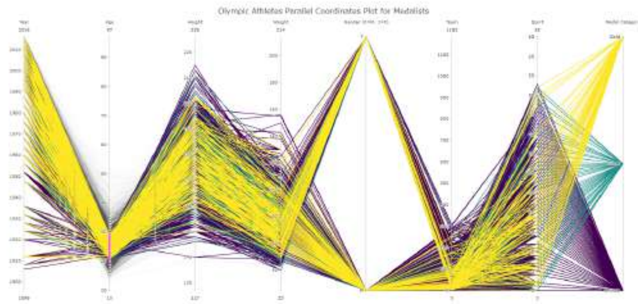


Fig. 64. Parallel Coordinates Plot for Olympic Dataset after brushing

3) Objective and Significance: This PCP visualization allows for multi-faceted exploration of athletic attributes associated with Olympic success:

- Analyze physical attributes that correlate with specific sports or medal outcomes.
- Examine age and gender distributions across sports to identify any demographic trends in medal achievement.
- Identify team and country patterns in specific sports, revealing trends that may indicate training or talent specialization within certain regions.

The interactivity of this PCP enables dynamic analysis, facilitating exploration of correlations that contribute to athletic performance and success at the Olympics.

4) Inferences and Observations: The Olympic Athletes PCP allows us to infer several detailed insights:

- **Physical Attributes by Sport**: Brushing along **Height** and **Weight** reveals sport-specific physical profiles among medalists. For instance, taller athletes appear more frequently in sports like basketball, while lighter, more agile athletes dominate gymnastics.
- **Age and Medal Success**: Axis reordering and brushing along **Age** reveal trends in age for different sports and medal types, suggesting that some sports favor younger athletes (e.g., gymnastics) while others may have older medalists (e.g., equestrian events).
- **Gender and Sport Patterns**: Brushing along the **Gender** axis highlights gender-based differences in participation and medal achievement in specific sports, which may point to gender-related physical demands or participation trends.
- **Country-Specific Trends**: By analyzing team codes alongside **Sport**, we can identify countries with strong performance in particular sports, hinting at specialized training programs or cultural affinity toward those sports.

5) Complexities and Disadvantages: Despite its insights, the Olympic Athletes PCP also presents interpretation challenges:

- **High Dimensionality**: With numerous sports, teams, and categorical variables, the plot can become visually complex, making it challenging to draw conclusions without targeted brushing.
- **Categorical Code Interpretation**: Like in the Jazz Network PCP, categorical codes (e.g., for **Team** and **Sport**)

lack inherent meaning, requiring users to cross-reference codes with specific labels, which may disrupt analysis flow.

- **Overlapping Data Points**: As with many athletes sharing similar attributes, lines may overlap, making it difficult to isolate individual data points without extensive brushing.

6) Conclusion: The Olympic Athletes PCP, with its brushing and axis reordering, provides valuable insights into the characteristics and demographic trends associated with Olympic success. While this interactivity enables detailed exploration, challenges such as high dimensionality and the interpretation of categorical codes underscore the need for careful analysis. Overall, the PCP offers a rich tool for exploring complex relationships in Olympic data, despite the interpretative challenges presented by overlapping data and categorical mappings.

D. Final Conclusions

Interactive PCPs offer a comprehensive and adaptable approach to analyzing complex datasets. Through the Jazz Network PCP, we identify influential musicians and collaboration patterns, while the Olympic Athletes PCP reveals physical and demographic characteristics associated with Olympic success. Brushing and axis reordering enhance the usability of PCPs, allowing for dynamic exploration, although challenges such as line overlap, categorical code interpretation, and dimensional complexity must be carefully managed to maintain clarity and insightfulness.

V. Author Contributions

A. Task 1 By Areen Vaghasiya

- Color maps, Tree maps

B. Task 2 By Aryan Vaghasiya

- Node Link diagrams, Contour maps

C. Task 3 By Rutul Patel

- Parallel Coordinate Plots, Quiver Plots and StreamLine Plots

References

[1] https://www.northwestknowledge.net/metdata/data/
[2] 120 years of Olympic data.
[3] https://gephi.org/datasets/jazz.net.zip