DAS732: Data Visualisation Assignment 2 Report

Aditya Saraf
IMT2022067
Aditya.Saraf@iiitb.ac.in

I. Introduction

This report examines scientific and information visualization techniques to create dynamic, insightful representations of real-world datasets. The project involves three key tasks:

- Scientific Visualization (SciVis): Using Python, we apply color mapping, contour mapping, and quiver plots to a selected three-month period of the gridMET dataset, capturing temporal variations in sea surface data. Experiments include diverse colormaps and quiver streamlining techniques to depict wind patterns.
- Network Visualization: We use Gephi [9] to generate node-link diagrams for a complex network, comparing three layout algorithms to highlight structural insights.
- Information Visualization (InfoVis): In Javascript and Python using Ploty.js [10], we create interactive parallel coordinates plots and treemaps, enabling multi-dimensional data exploration through axis reordering and partitioning strategies.

Each section outlines the methods, key findings, and visual inferences drawn, showcasing effective use of SciVis and InfoVis techniques for data interpretation.

II. Dataset

The dataset comprises several files detailing attrition rates across various demographics and geographical areas. Key variables include:

1) gridMET Dataset [1]: Comprising several files that detail daily meteorological data across the contiguous United States for the period of July-September 2019, used for scientific visualization (SciVis).
2) hero-social-network.gephi [2]: A network dataset utilized for network visualization.
3) HE Attrition Data Main (2005-2013) [3]: This dataset contains attrition rates across various demographics and geographical areas, employed for information visualization (InfoVis).

III. Tasks

Through visual exploratory analysis, we aim to gain the following insights and expect one to reproduce the following tasks:

1) T1: Scientific Visualization (SciVis)
2) T2: Detailed Environmental Analysis (SciVis)
3) T3: Network Visualization of Marvel Social Network (InfoVis)
4) T4: Parallel Coordinate Plot and TreeMap Analysis (InfoVis)

IV. Assumption / Data Filtration

Given the datasets, the necessary filtration was performed using np [4] for Scivis and data was read using xarray [6]. For Infovis, node filtration was applied based on node degrees for the node-link diagram, and for the treemap, empty attribute values were replaced with 0.

V. Data Stories

A. *T1: Scientific Visualization (SciVis)*

1) Reason to select this timeline: The chosen timeline of July to September 2019, for the gridMET dataset is particularly significant, as it encompasses peak summer months when the United States often experiences critical climate events. This period includes instances of major wildfires, severe drought conditions, and elevated temperatures that notably affect various regions, particularly in July, which is often one of the hottest months. Additionally, the timeline allows for the examination of hurricane impacts, which can occur in summer, providing a comprehensive view of weather extremes during this period. Such data offers valuable insights into the effects of climate variability on natural disasters, crucial for scientific visualization and analysis.

2) Introduction: Wind speed and direction data are essential for many fields such as meteorology and environmental studies. Quiver and streamline plots provide effective visualization tools for vector fields, helping to represent the strength and direction of winds across different regions. This report describes various types of quiver and streamline plots used to visualize wind data, emphasizing their distinct features and configuration.

- Quiver Plots
  Quiver plots are used to illustrate vector fields with arrows pointing in the direction of flow, where arrow length can indicate vector magnitude. This section provides examples of different quiver plot configurations.
  – Black Quiver Plot : In this quiver plot (Figure 1), all arrows are uniformly colored black. This type of plot is effective for representing direction and magnitude where smaller length shows lower value in magnitude and vice versa.

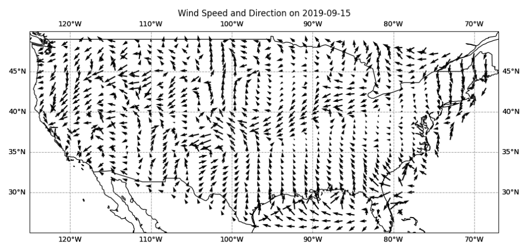

Fig. 1: Quiver plot with black arrows showing wind direction.

– Quiver Plot with Continuous Color Map : In this version (Figure 2), arrow color varies continuously based on the wind speed. A color gradient (e.g., plasma) helps visualize changes in magnitude across the field.

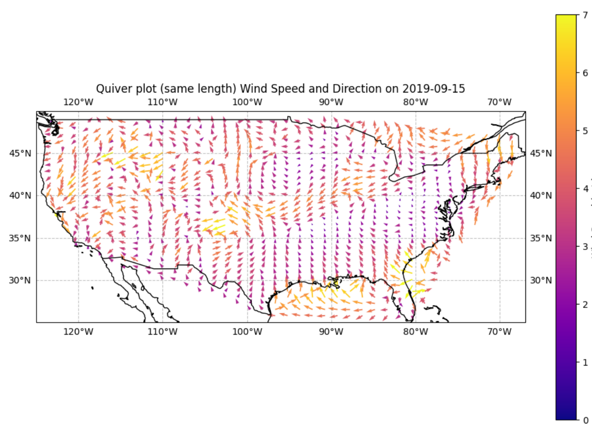

Fig. 2: Quiver plot with continuous color map representing wind speed.

– Quiver Plot with Discrete Colors : Here (Figure 3), arrow colors are mapped to discrete categories, each representing a range of speeds. This approach is useful when speed data is classified into bins, like low, medium, and high speeds.

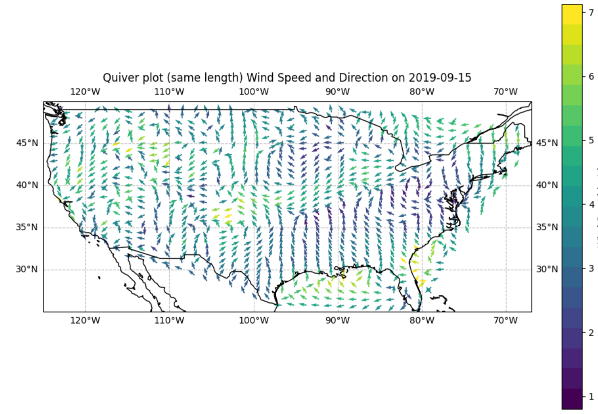

Fig. 3: Quiver plot with discrete colors indicating different speed ranges.

– Scaled Quiver Plot: This quiver plot (Figure 4) type scales for longitude and latitude for creation for grid.

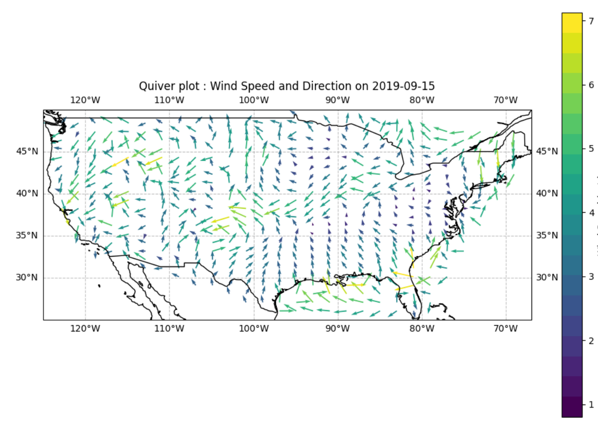

Fig. 4: Scaled quiver plot with same arrow length with viridis colormap.

- Streamline Plots
  Streamline plots visualize flow as continuous lines, representing the path of a particle in the flow field. These plots provide a smooth, intuitive sense of direction and magnitude over a vector field.
  – Streamline Plot with Viridis Color Map : This plot (Figure 5) uses the viridis color map, with line colors varying based on speed. The continuous color gradient makes it easy to identify areas of higher or lower speeds in the flow field.

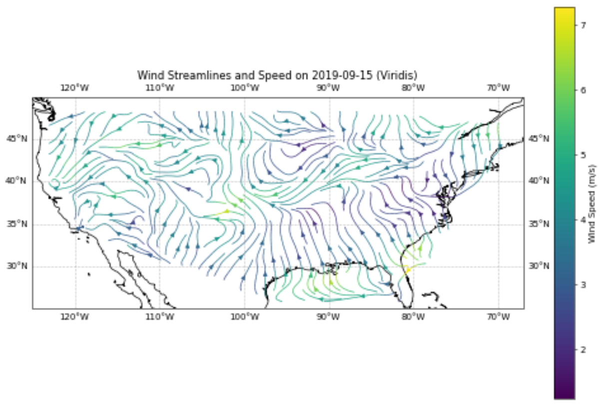

Fig. 5: Streamline plot with viridis color map showing speed variations.

– Streamline Plot with Dark Blue Color : In this version (Figure 6), streamlines are uniformly colored dark blue, focusing solely on the direction of flow without representing speed variations. This is effective for representing flow structure clearly and uniformly.

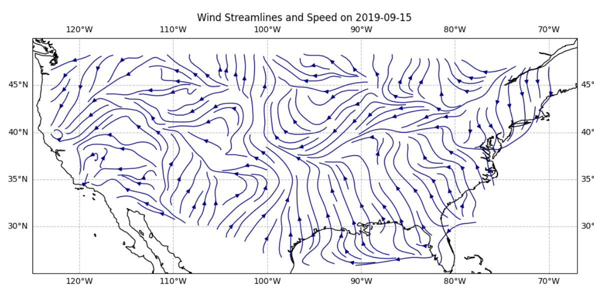

Fig. 6: Streamline plot with dark blue color showing flow direction.

3) Conclusion: The choice of quiver or streamline plot style depends on the visualization objective. Black or dark blue colors are effective for representing direction alone, while color-mapped and scaled versions convey both speed and direction. Streamline plots are particularly useful for illustrating continuous flow and provide a more intuitive understanding of wind patterns.

B. T2: Detailed Environmental Analysis (SciVis)

This subsection of the report explores the role of environmental factors in extreme weather events, focusing specifically on wildfires, droughts, and hurricanes. Through visualizing key variables such as temperature, fuel moisture, and vapor pressure deficit, the analysis sheds light on wildfire dynamics and their relationship with environmental conditions. Additionally, the project examines drought severity through indices like the Palmer Drought Severity Index and investigates hurricane patterns, using data such as wind speed and precipitation levels to understand their formation and intensity. The aim is to provide insights into the interconnection between these weather events and their environmental drivers through clear and effective visual representations created gifs using image.io [7] and pillow [8].

1) Major Wildfires: Cow Wildfire Analysis: This analysis focuses on significant wildfire event—the Cow fires—examined through a series of spatially zoomed maps centered on the affected areas. By narrowing the view to these specific regions, we aim to capture detailed environmental conditions influencing fire's progression. Both fires were initially ignited by lightning and were further fueled by dry atmospheric conditions and high wind speeds.

Cow Wildfire The Cow Fire, originating in a remote region of Northern California, began on July 20, 2019. By the time it was contained in early September, the fire had scorched approximately 15,000 acres of forested land, with an estimated damage cost of around $2.5 million. Key factors contributing to the fire's spread included elevated vapor pressure deficits (VPD) and strong wind speeds.

The wildfire conditions are visualized using:

- Wind Speed and Direction: Represented by quiver plots (black colour), wind vectors highlight the direction and speed, critical in understanding fire spread. Higher wind speeds and consistent directions often exacerbate wildfire progression, especially when aligned with dry weather conditions.
- Vapor Pressure Deficit (VPD): Shown through contour plots (colormap: discrete OrRd), VPD is a measure of the drying power of the atmosphere and strongly correlates with fire risk. High VPD values indicate drier air, which facilitates moisture loss in fuels, making them more prone to ignition.
- Dead Fuel Moisture Content: Presented as a heatmap (colormap: continuous Greens), this parameter shows the moisture in dead vegetation, which becomes highly combustible under dry conditions. Lower moisture levels indicate a greater likelihood of sustained fire spread.

The initial locations of wildfire are marked distinctly in the visualizations. For the Cow Fire, the area of origin is highlighted in blue on July 20, indicating the starting points for wildfire.

Cow Wildfire (July 20 - September 1, 2019)

Figures 7 to 12 detail the progression of the Cow Fire. On July 20, 2019, the wildfire began in an area highlighted in blue, with moderate VPD and low fuel moisture, as seen in Figure 7. The following days show an increase in VPD (deeper red in contour plots) and a decrease in fuel moisture (darker green in heatmaps), contributing to intensified fire conditions, as in Figures 9 and 11. By September 1 (Figure 12), fuel moisture levels began to recover, coinciding with reduced fire activity.

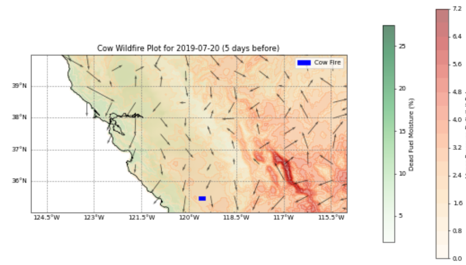

Fig. 7: Cow wildfire condition on 2019-07-20

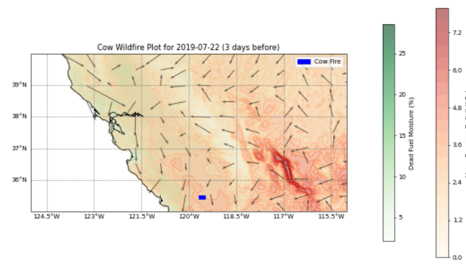

Fig. 8: Cow wildfire condition on 2019-07-22

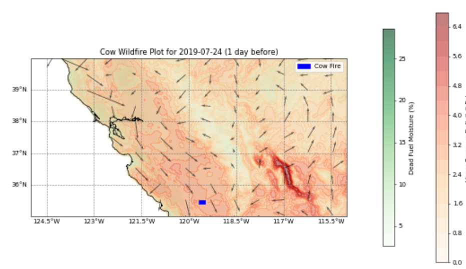

Fig. 9: Cow wildfire condition on 2019-07-24

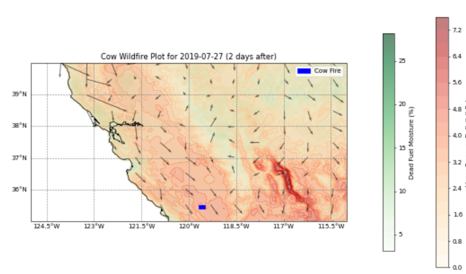

Fig. 10: Cow wildfire condition on 2019-07-27

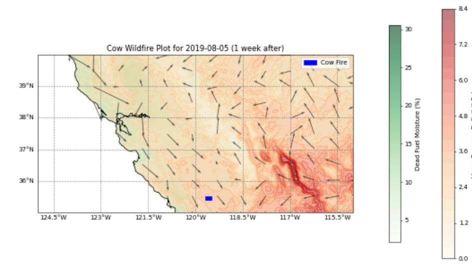

Fig. 11: Cow wildfire condition on 2019-08-05

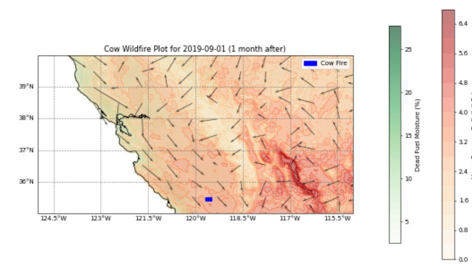

Fig. 12: Cow wildfire condition on 2019-09-01

Conclusion: These figures underscore the strong impact of meteorological conditions on wildfire progression. High VPD and low fuel moisture levels, combined with strong and directional wind, amplified the spread of both the Cow wildfire. By illustrating these factors, this analysis provides insights into the conditions under which wildfires evolve, aiding in future predictions and preparedness measures.

2) Overall Drought Conditions: This analysis explores drought conditions across a wide region in the United States from July to September 2019, a period marked by significant drought events.

Each figure illustrates conditions on a specific date, capturing the combined impact of wind direction, Palmer Drought Severity Index (PDSI), and average temperature levels across the region. Notably, areas with inland wind sources, which lack moisture from the ocean, tended to experience more severe drought effects.

- Wind Speed and Direction: Represented by dark blue streamline plots, the wind vectors indicate direction and intensity. This approach is effective for visualizing wind patterns, allowing for an intuitive understanding of airflow without cluttering the map, which is essential for interpreting the impact of wind in spreading or alleviating drought conditions.
- Palmer Drought Severity Index (PDSI): Displayed as a continuous heatmap using a diverging BrBG colormap, PDSI values highlight areas with moisture deficits or surpluses. Lower PDSI values (in shades of brown)

indicate more severe drought conditions, while higher values (in green) suggest wetter areas. The BrBG colormap was chosen for its ability to represent the severity spectrum effectively, where brown tones intuitively signal dryness, enhancing comprehension.

- Average Temperature: Shown with a discrete contour plot using a coolwarm colormap, this variable indicates temperature variations. Warmer colors (reds) represent high temperatures, while cooler colors (blues) depict lower temperatures. Contour plots enable a clear demarcation of temperature bands across regions, aiding in understanding how temperature influences drought severity.

Impact and Economic Loss: The severe drought during this period led to substantial agricultural losses, particularly in the Midwest and Southwest regions. Crop failures were common in these areas, and the economic impacts are estimated at approximately $1.5 billion, driven by increased costs for feed and water scarcity affecting livestock.

Explanation of Plot Choice and Color Scheme: The choice of streamline plots for wind direction provides a dynamic, uncluttered view of airflow patterns, while heatmaps for PDSI allow continuous representation of drought severity. Contour plots effectively delineate temperature gradients, assisting in identifying hotspots. The selected colormaps—BrBG for PDSI and coolwarm for temperature—enhance the visual differentiation between dry and wet conditions and between high and low temperatures, respectively, aligning with intuitive color associations for temperature and moisture.

The following figures show the progression of drought severity over time. These visuals emphasize regions with persistently low PDSI values, high temperatures, and inland wind patterns, resulting in severe, sustained drought impacts.

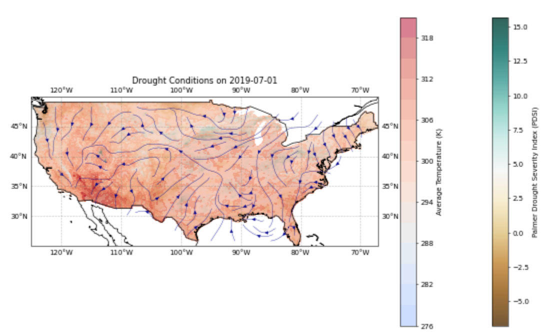

Fig. 13: Drought conditions on 2019-07-01

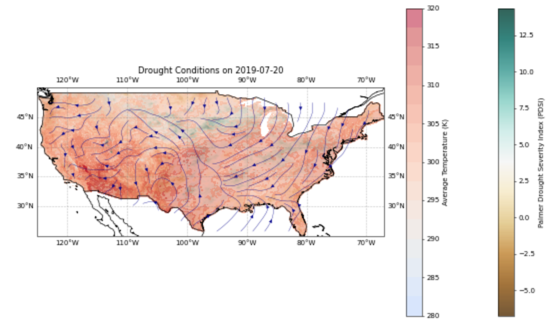

Fig. 14: Drought conditions on 2019-07-20

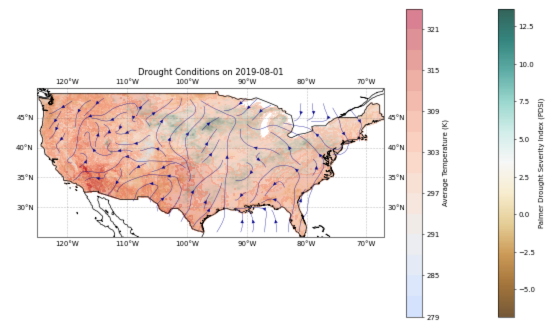

Fig. 15: Drought conditions on 2019-08-01

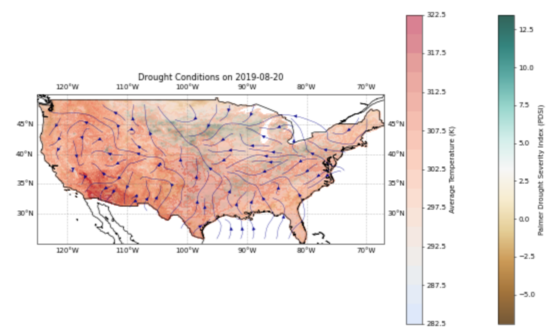

Fig. 16: Drought conditions on 2019-08-20

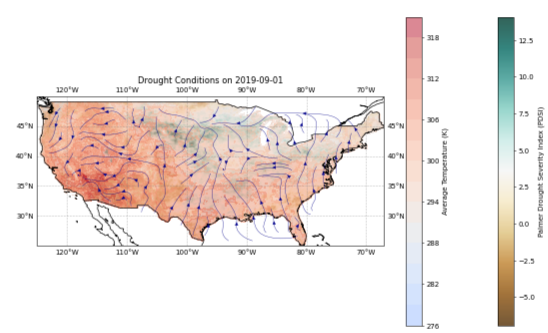

Fig. 17: Drought conditions on 2019-09-01

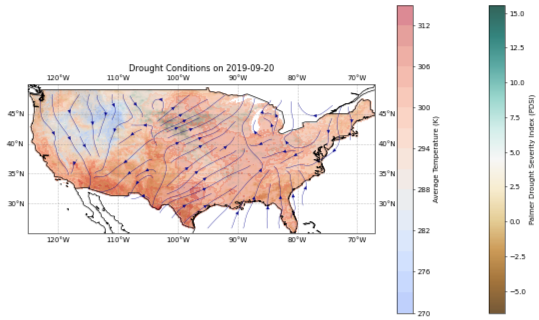

Fig. 18: Drought conditions on 2019-09-20

Conclusion: This analysis highlights the influence of temperature, moisture availability, and wind direction on drought severity. Regions with consistently low PDSI values, high temperatures, and lack of ocean-derived wind influence experienced the most severe drought impacts, underscoring the interaction of these variables in drought dynamics.

3) Hurricane Barry: This section examines Hurricane Barry, which impacted the Gulf Coast in July 2019, causing significant flooding and wind damage. The figures below display Hurricane Barry's conditions over time, with each figure highlighting a different day of the storm's progression from July 10 to July 15, 2019. An elliptical overlay, shown in varying colors, represents the primary impact zone and indicates storm intensity within this region. Variable Representations and Color Schemes:

- Wind Speed and Direction: Shown with dark blue streamline plots, wind vectors depict both direction and intensity, emphasizing rotational patterns and wind convergence associated with the storm's core.
- Precipitation (mm): Represented by a discrete contourf plot with the ”Blues” color scheme, where darker blues indicate higher rainfall intensities, highlighting areas vulnerable to flooding.
- Specific Humidity: Visualized using a continuous ”YlGnBu” heatmap, specific humidity values reflect the moisture content in the atmosphere. Higher values appear in green to yellow shades, signaling areas with significant atmospheric moisture.

Progression and Visual Changes: Each figure illustrates how wind, precipitation, and humidity conditions evolved during Hurricane Barry's approach and landfall:

- July 10 (Fig. 19) – Initial conditions show moderate humidity and precipitation levels in the southern coastal areas.
- July 11 (Fig. 20) – Humidity increases, and precipitation areas expand, indicating the storm's strengthening.
- July 12 (Fig. 21) – The elliptical impact area turns red, showing intensified storm impact and higher wind activity in the core region.
- July 13 (Fig. 22) – The impact zone remains red, with widespread precipitation and high humidity, signaling the peak of the storm.
- July 14 (Fig. 23) – The elliptical zone shifts to a purple shade, suggesting a reduction in storm intensity as wind speeds and precipitation begin to decrease.
- July 15 (Fig. 24) – The impact area shows lighter colors, with wind patterns dispersing, indicating the storm's weakening and dissipation.

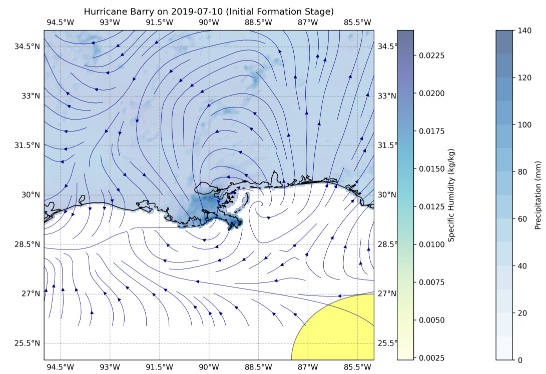

Fig. 19: Hurricane Barry conditions on 2019-07-10

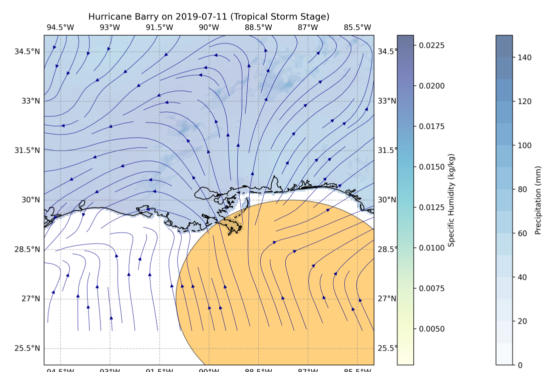

Fig. 20: Hurricane Barry conditions on 2019-07-11

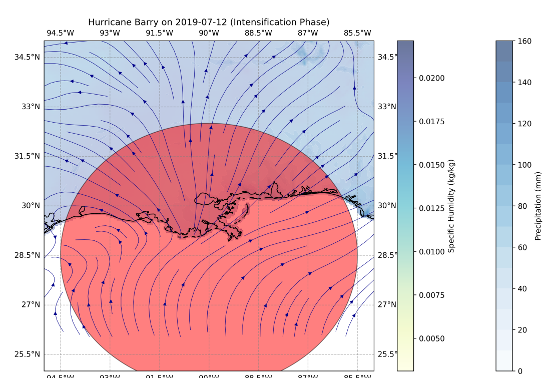

Fig. 21: Hurricane Barry conditions on 2019-07-12

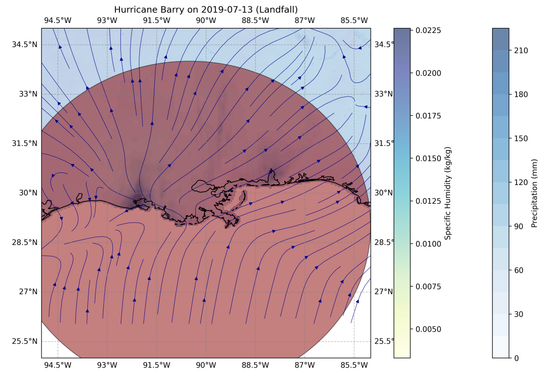

Fig. 22: Hurricane Barry conditions on 2019-07-13

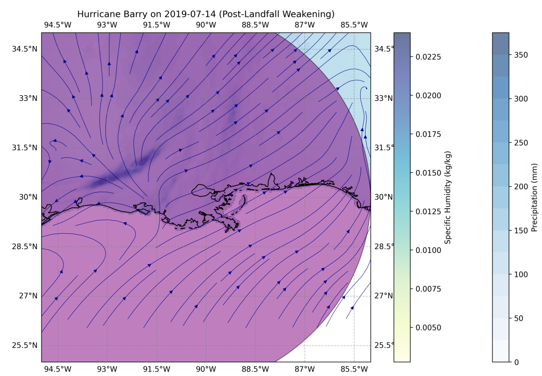

Fig. 23: Hurricane Barry conditions on 2019-07-14

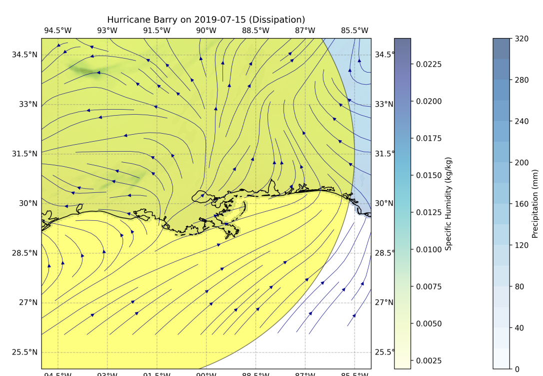

Fig. 24: Hurricane Barry conditions on 2019-07-15

Impact and Economic Loss: Hurricane Barry resulted in significant flooding and wind damage, especially in Louisiana and Mississippi, with estimated economic losses over $500 million.

Conclusion: This analysis highlights the interaction of wind, humidity, and precipitation, helping to identify the regions that experienced the most severe impacts. The color scheme and elliptical overlay effectively convey changes in storm intensity and distribution across affected areas.

C. T2: Network Visualization of Marvel Social Network (InfoVis)

The Marvel Social Network dataset consists of a simple representation of relationships among characters in the Marvel universe. It has connections between characters, weighted by number of common appearances in comics.

Key features of the dataset include:

- Nodes: Each node represents a Marvel character, identified by their names.
- Edges: The dataset features undirected weighted edges that indicate the relationships between characters, where each weight signifies the frequency of shared comic appearances. An edge between two nodes implies that the connected characters have appeared together in the same comic, highlighting alliances, friendships, or frequent collaborations.

Note: To enhance visualization and engage enthusiasts, We have added images for major nodes, labeled under the column name “image.” This visual element enriches the representation and fosters a deeper connection to the characters.

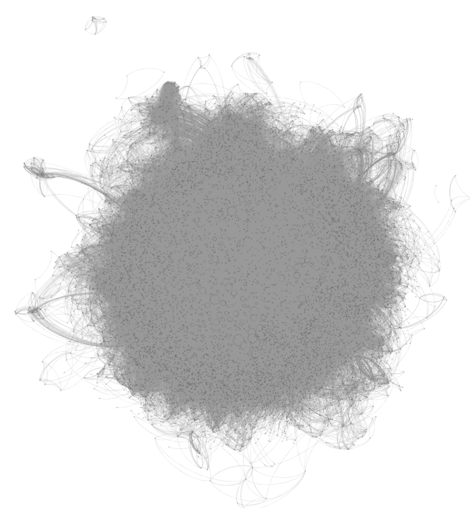

Fig. 25: Original network graph of Marvel social network

The dataset initially comprises 10,469 nodes and 178,115 undirected edges and looks like as shown in Figure 25. This dataset is essential for network visualization, enabling the exploration of social dynamics within the Marvel universe. Analyzing these connections allows us to uncover patterns, identify influential characters, and visualize the overall network structure.

The network consists of major nodes, represented in shades of green as shown in Figure 26. Darker shades indicate nodes with a higher degree of connectivity, while lighter shades represent those with fewer connections. The treemap (Figure 48), in particular, allows us to verify the relative importance of these nodes by their edge weight distribution.

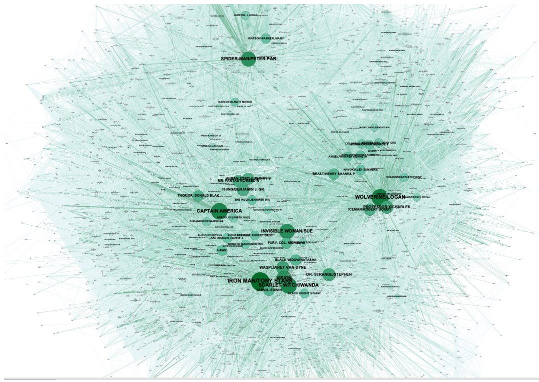

Fig. 26: Major Nodes for Marvel Social Network

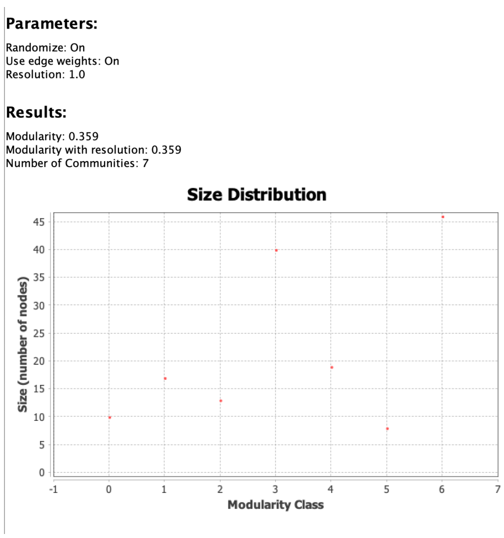

Fig. 27: Modularity of filtered nodes

To enhance network visualization, nodes were filtered based on degree, with an initial threshold of greater than 531. Additional filtering at a threshold of greater than 335 (Figure 27 shows modularity for communities) was applied to reveal more clusters.

1) Color Scheme for Network Visualization: We opted for a costume-inspired color scheme for major nodes in each cluster, drawing from the iconic colors of key Marvel heroes to reflect their centrality and influence in the Marvel Social Network. This color scheme was designed based on node modularity and betweenness properties, with the most connected characters represented by distinctive colors.:

- Yellow: Represents Wolverine from the X-Men, symbolizing connections to various mutants and heroes.
- Light Blue: Represents characters from the Fantastic Four, emphasizing alliances and interactions unique to this team.
- Red: Represents Avengers like Iron Man, Captain America, and Scarlet Witch, illustrating their central roles in broader Marvel alliances.
- Dark Blue: Represents Spider-Man, inspired by his classic red and blue suit, highlighting his wide-reaching connections across teams and solo interactions.
- Light Red: Represents other Marvel heroes, including Doctor Strange and Hulk, highlighting their connections across various teams without belonging to a specific alliance.
- Grey: Used for Asgardians, symbolizing their characteristic armor and metallic attire, which is distinctive within the network.
- Green: Applies to other heroes, providing a distinct visual cue for characters who do not fall into major team categories but still hold connections within the Marvel universe.

Note: The nodes with images are outlined with specific border colors as mentioned above, which indicate their respective communities. This visual distinction enhances the clarity of community affiliations within the Marvel social network.

2) Different Layouts: We explored various layout algorithms for visualizing the Marvel Social Network, ultimately selecting ForceAtlas, OpenOrd, and Fruchterman-Reingold as the most effective. Each of these layouts was chosen for its ability to clearly represent the network structure and enhance the exploration of relationships among characters.

1) ForceAtlas Layout: This layout is known for its effectiveness in representing social networks by simulating physical forces, where nodes repel each other while edges act as springs. This results in a clear visual separation of densely connected clusters.
   Purpose: The ForceAtlas layout uses a physics-based approach, simulating repulsive and attractive forces between nodes to create a balanced and visually appealing cluster arrangement. This layout emphasizes clusters and makes it easy to spot main groups and alliances.
   Execution Steps:
   - Edge Weight Influence: Adjusted to highlight strong connections based on character co-appearances.
   - Repulsion Strength: Set to a moderate level to prevent nodes from clustering too densely.
   - No Overlap and Label Adjust: Enabled no overlap and label adjust features to ensure node and label clarity within each cluster.

   The following visualizations were created:
   - Proportional Node Size Visualization (Figure 28): Node sizes correspond to their degree, allowing for the identification of influential characters.

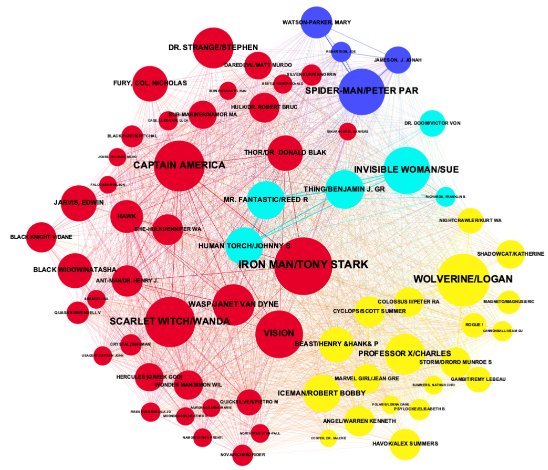

Fig. 28: ForceAtlas Layout: Filtered

- Node Size with Images and Labels (Figure 29): This version enhances engagement by incorporating character images and labels, maintaining node size based on degree.

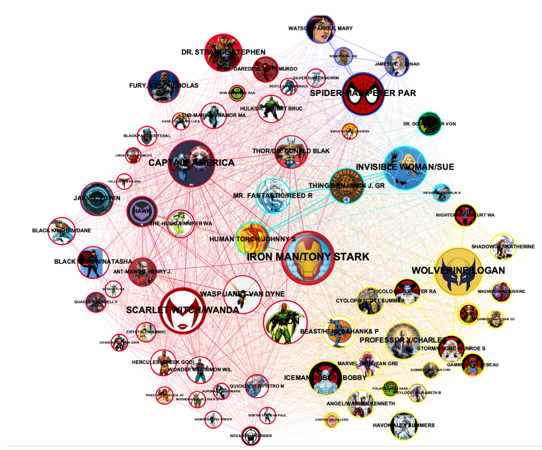

Fig. 29: ForceAtlas Layout with images: Filtered

Outcome: The ForceAtlas layout effectively highlighted major clusters like the Avengers and X-Men. Prominent characters appeared more central, while the layout's natural force simulation allowed a clean separation between groups.

2) Fruchterman-Reingold layout: It uses a force-directed algorithm to balance forces among nodes, creating a visually clear and balanced network structure. It highlights clusters and relationships effectively, enhancing readability and aesthetic appeal.: This layout is designed for large networks and aims to provide an aesthetic representation of connections by minimizing edge crossings. Execution Steps:

- Area Factor Adjustment: Increased to spread out nodes and reduce density.
- Convergence Speed: Set to a higher value for faster rendering without losing node distribution quality.
- No Overlap and Label Adjust: Enabled to prevent node overlap and improve label readability.

The following visualizations were created:

- Proportional Node Size Visualization (Figure 30): Node sizes correspond to their degree, allowing for the identification of influential characters.

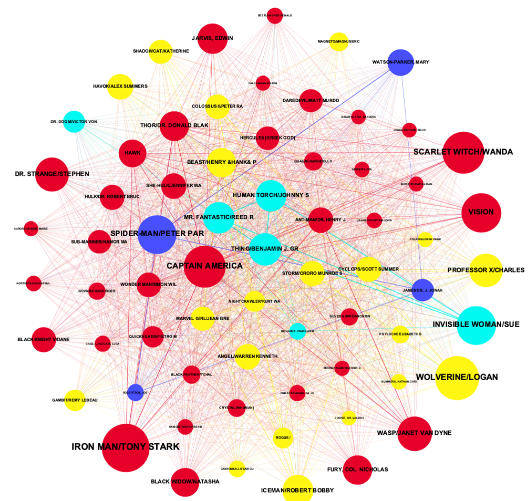

Fig. 30: Fruchterman-Reingold Layout: Filtered

- Node Size with Images and Labels (Figure 31): This version enhances engagement by incorporating character images and labels, maintaining node size based on degree.

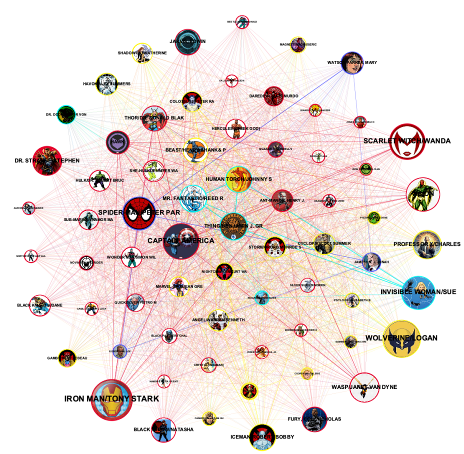

Fig. 31: Fruchterman-Reingold Layout with images: Filtered

Outcome: This layout produced well-spaced clusters with minimal edge overlap. It maintained clear divisions among Marvel teams, allowing a readable structure with both central and peripheral nodes visible.

3) OpenOrd Layout: This layout is effective for highlighting community structures and improving overall readability in dense networks. Execution Steps:
   - Edge-Cutting Phase: Enhanced to cut weak connections, clarifying boundaries between distinct character alliances.
   - Expansion and Cooldown Phases: Increased the duration to refine the layout and form clear clusters.
   - Local Optimization: Enabled local optimization to create tighter clusters for groups like minor teams or recurring alliances.
   - No Overlap and Label Adjust: Applied to further enhance readability within each distinct cluster.

   The following approaches were taken:
   - Proportional Node Size Visualization (Figure 32): Node sizes correspond to their degree, allowing for the identification of influential characters.

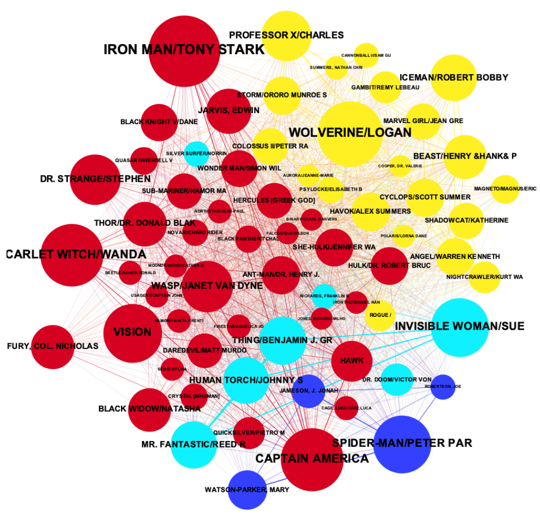

Fig. 32: OpenOrd Layout: Filtered

   - Node Size with Images and Labels (Figure 33): This version enhances engagement by incorporating character images and labels, maintaining node size based on degree.

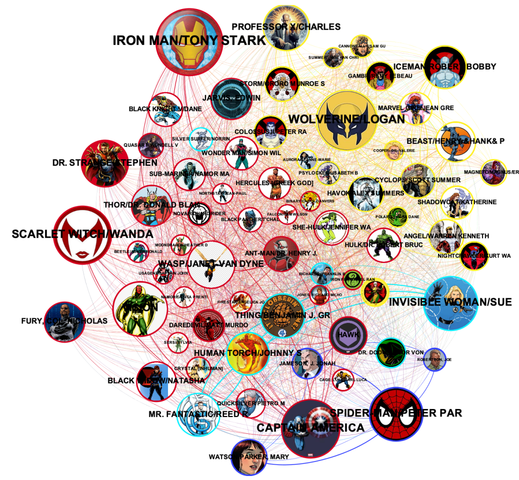

Fig. 33: OpenOrd Layout with images: Filtered

Outcome: The OpenOrd layout provided a visually effective depiction of secondary relationships within the Marvel universe. Character clusters formed more visibly, illustrating subtle alliances and subgroupings within larger teams.

.

4) Additionally, nodes with a degree > 335 were filtered to capture smaller but significant clusters within the network. This provides a more detailed view of sub-communities, as depicted in Figures 34 35 36. This approach allows us to observe secondary relationships and connections among Marvel characters, complementing the main visualization.

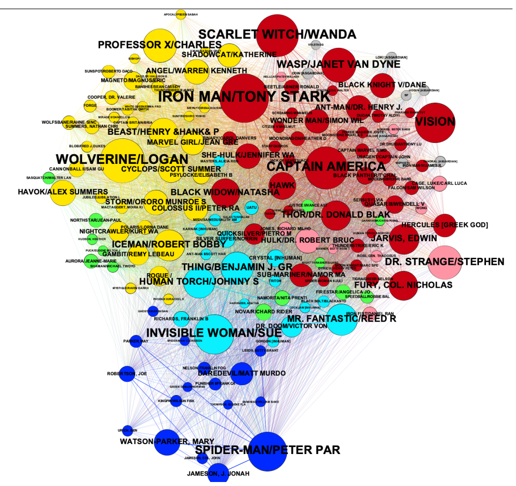

Fig. 34: OpenOrd Layout: Nodes Filtered by Degree > 335

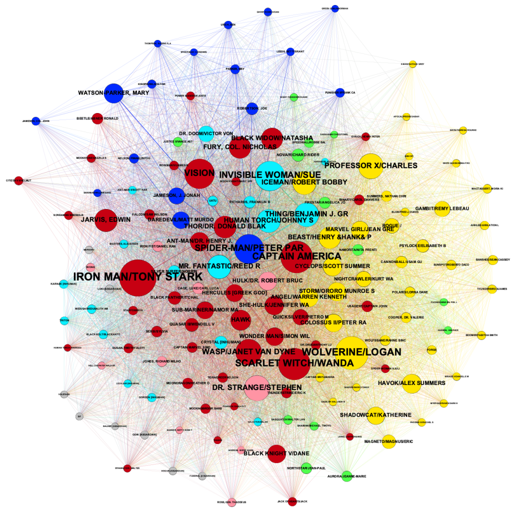

Fig. 35: Fruchterman-Reingold Layout: Nodes Filtered by Degree > 335

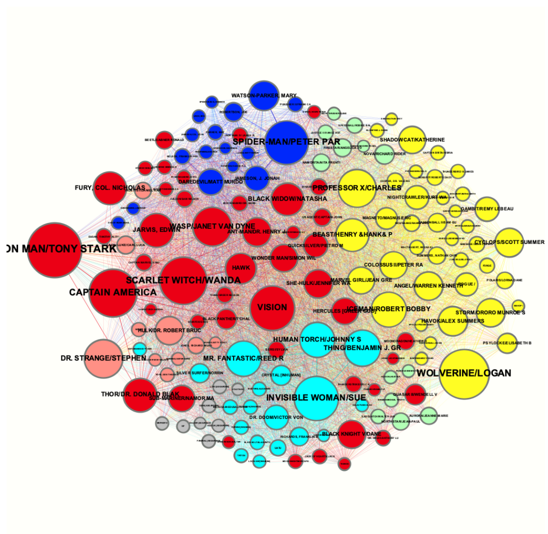

Fig. 36: ForceAtlas Layout: Nodes Filtered by Degree > 335

5) Additional Techniques and Metrics: For all layouts, **modularity** and **betweenness centrality** metrics were applied to detect and color unique groups. Modularity helped in identifying distinct clusters, while betweenness centrality highlighted the most influential nodes, allowing for clearer visualization of key characters and connections. Each group was color-coded to emphasize narrative alliances, improving the interpretability of the network

3) Inferences: From these plots, the following key observations are made:

- Central Characters: The visualizations reveal that characters such as Iron Man, Wolverine, Captain America, Invisible Woman, and Spider-Man are among the most connected within the network. These characters, having high degree values, act as central figures, influencing numerous storylines and character interactions.
- Community Structure: The network exhibits distinct clusters, particularly evident among teams like the Avengers, X-Men, Fantastic Four, Spiderverse, Hulk & Dr. Strange Community, Asgardians and other community. These groups are characterized by dense interconnections, highlighting strong relationships that suggest potential collaboration and narrative arcs shared within these teams.
- Visual Patterns: The color coding of nodes, based on character affiliations, aids in quickly identifying major groups. For example, characters from the Avengers are marked in red, while X-Men characters are represented in yellow. This thematic representation enhances the clarity of the network's structure.
- Impact of Filters: Applying filters based on node degree significantly alters the visualizations. Focusing on characters with degrees greater than 531 reveals key players in the network, while filtering those above 336 uncovers additional clusters and less prominent characters, providing a deeper understanding of the overall landscape.
- Thematic Representation: The visualized relationships illustrate broader themes present in the Marvel narrative, such as heroism and teamwork. The clustering of characters around shared colors and sizes reflects their collaborative spirit and narrative significance.
- The ForceAtlas layout outperforms the other layouts by effectively highlighting the formation of distinct clusters within the Marvel Social Network, allowing for clearer visualization of relationships among different character groups.

4) Summary: The Marvel Social Network dataset visualizes relationships among characters, weighted by their comic appearances, to explore social dynamics within the Marvel universe. It includes 10,469 nodes (characters) and 178,115 edges (connections), where major characters are highlighted with images. Filtering by node degree reveals significant clusters and influential characters. The network uses a color scheme inspired by iconic costumes to distinguish affiliations—red for Avengers, yellow for X-Men, light blue for Fantastic Four, etc. Visualizations through ForceAtlas, Fruchterman-Reingold, and OpenOrd layouts showcase character clusters, with ForceAtlas best revealing distinct groupings. Key observations include central characters like Iron Man and Wolverine, community structures (e.g., Avengers, X-Men), and thematic representations reflecting Marvel's collaborative themes.

D. T4: Parallel Coordinate Plot and TreeMap Analysis (InfoVis)

1) Parallel Coordinate Plot: The Parallel Coordinate Plot (PCP) visualizations below illustrate the variation in student attrition rates across different attributes. PCP's are particularly effective for multidimensional datasets, enabling us to observe trends, clusters, and correlations across multiple attributes simultaneously.

Figure Descriptions and Analysis

- Overall Higher Attrition Rates by Attribute in Australia (Figure 37): This plot visualizes attrition rates across different years with variety of attributes, such as socio-economic status, indigenous status, and study mode, providing a broad view of student retention patterns in Australia. The lines represent individual data points across attributes, revealing trends and potential relationships among them.

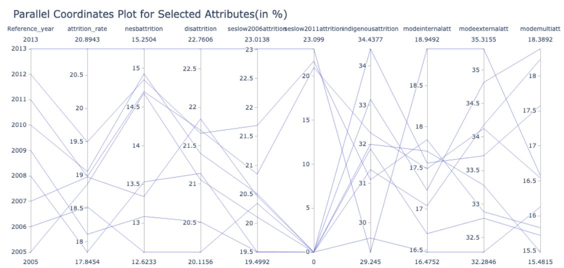

Fig. 37: Parallel Coordinate Plot for attrition rate based on different attributes

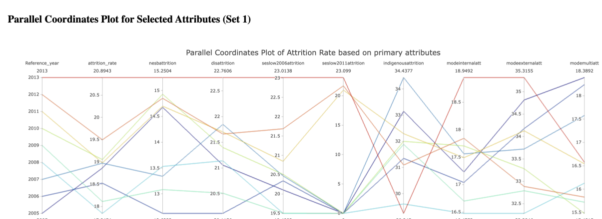

Fig. 38: Parallel Coordinate Plot (Jet colormap) for attrition rate based on different attributes

- Attrition Rates with Brushing by Specific Ranges (Figure 39): This PCP employs brushing to filter and highlight specific ranges, making it easier to focus on specific-attrition groups. It also reorders some axes. By isolating certain ranges, we can better understand the demographic factors associated with elevated attrition rates.

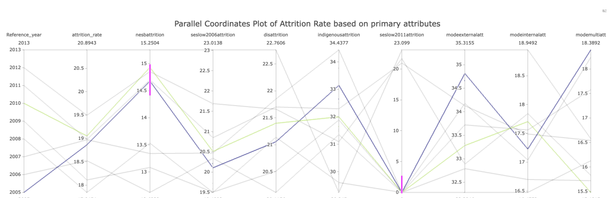

Fig. 39: Parallel Coordinate Plot: Brushing and Reordering

- Attrition Rate by Qualification Level(Figure 40): This plot focuses on attrition rates segmented by student qualification levels across Australian institutions. By visualizing qualification levels, we can observe how academic progression correlates with attrition trends.

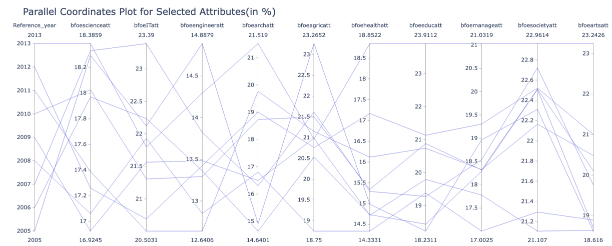

Fig. 40: Parallel Coordinate Plot for attrition rate based on qualification

- Attrition Rate by Broad Field of Education (BFOE) in Australia (Figure 41): Here, attrition rates are broken down by Broad Field of Education (BFOE), providing insight into how different academic disciplines affect student retention.

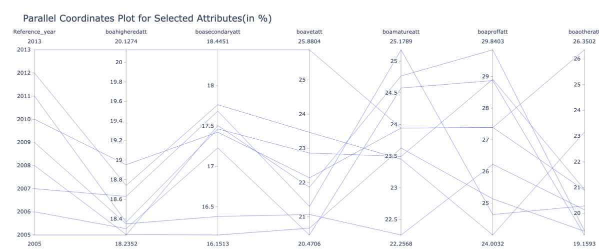

Fig. 41: Parallel Coordinate Plot for attrition rate based on Broad Field of Education

2) Tree Map Analysis: The treemap visualizations below provide a hierarchical view of student attrition data across different dimensions, helping us explore relationships within the dataset based on the attribute structure and temporal variations.

Figure Descriptions and Analysis:

- Hierarchy Based on Year → Attribute (Figure 42): This treemap organizes the data hierarchically by year at the top level, with each year's data further subdivided into different attributes related to student attrition. This structure allows us to compare how each attribute contributed to attrition rates within specific years.
  - Key Insight: By visualizing attrition data in this manner, we can quickly identify years with particularly high or low attrition rates across multiple attributes, helping to highlight any period-specific patterns or anomalies.
  - Color Insight: The use of red for high attrition rates quickly draws attention to years with significant student losses, allowing for an intuitive assessment of risk factors within each year.

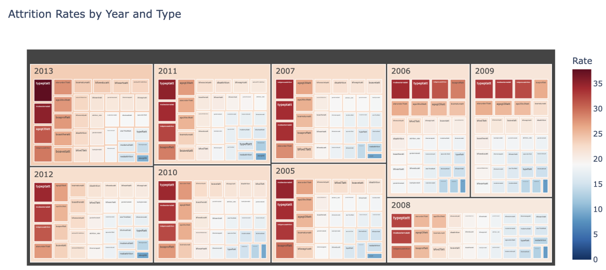

Fig. 42: TreeMap: Heirarchy based on Year → Attribute

- Hierarchy Based on Attribute → Year (Figure 43): Here, the hierarchy is reversed to organize the data by attribute at the top level, with each attribute containing data for multiple years. This layout allows us to observe the temporal trend within each attribute individually, providing a clear view of the attribute's performance over time.
  - Key Insight: This perspective shows the progression or stability of specific attributes across the years, helping identify whether certain factors (e.g., socio-economic background or mode of study) consistently contribute to attrition rates over time or if they vary significantly by year.
  - Color Insight: Red-shaded sections indicate attributes that consistently contribute to high attrition, while blue shades denote areas with lower attrition. This enables easy identification of long-term, attribute-based risks, facilitating targeted analysis of attributes.

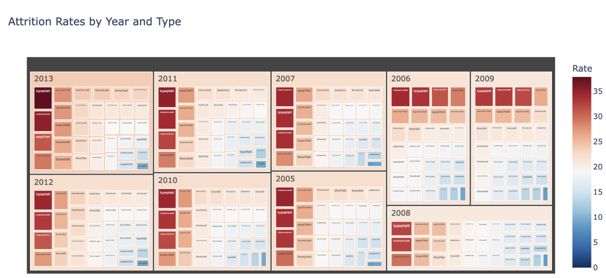

Fig. 43: TreeMap: Heirarchy based on Year → Attribute

- Different Spatial Partitioning Analysis: In this analysis, various treemap partitioning techniques are explored to assess how they represent data distribution visually. Treemaps provide a compact way to display hierarchical data, making it easier to identify patterns or outliers. Four different spatial partitioning techniques were used to create treemaps: Squarify, Binary, Dice, and Slice. Each approach offers a unique layout and emphasizes different aspects of the data structure.
  - Squarify: Prioritizes balanced square shapes, making it easier to compare relative sizes across categories.

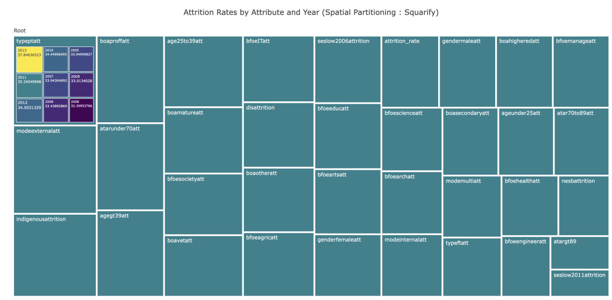

Fig. 44: TreeMap Based on Squarify

  - Binary: Alternates between horizontal and vertical splits, providing even space distribution, though less balanced in shape.

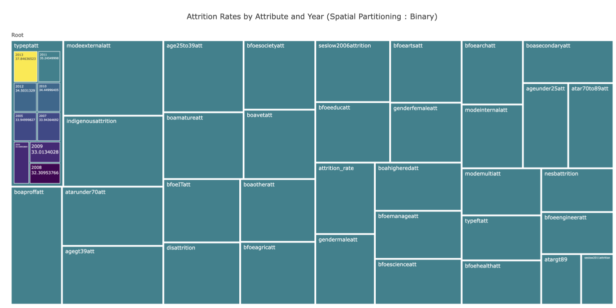

Fig. 45: TreeMap Based on Binary

  - Dice: Divides space into horizontal slices, ideal for simpler hierarchical structures, emphasizing proportions.

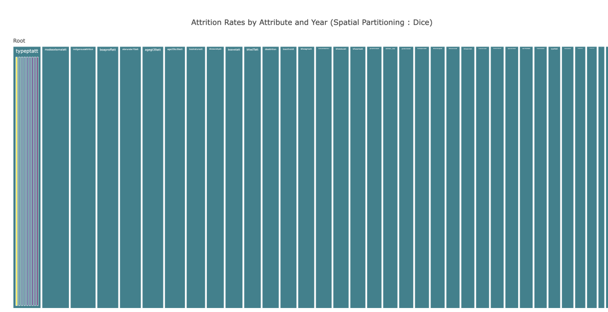

Fig. 46: TreeMap Based on Dice

  - Slice: Creates vertical slices, useful for comparing categories at a single hierarchy level.

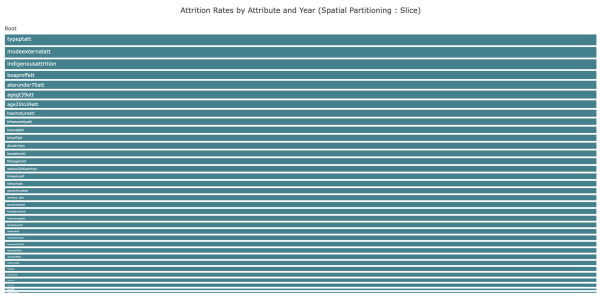

Fig. 47: TreeMap Based on Slice

Each technique highlights different aspects of the data, from balance (Squarify) to hierarchical layering (Dice and Slice).

- Hierarchy of Marvel Social Network Based on Character Connections (Figure 48): This treemap, based on a Node Link diagram dataset, visualizes the Marvel social network where nodes represent characters and edges denote relationships or interactions between them. The hierarchy is based on the number of connections (degrees) each character has, offering insights into character prominence and social clustering within the Marvel universe.
  - Key Insight: The hierarchical structure showcases major characters with numerous connections, highlighting their central role within the network. Characters with high connectivity likely serve as focal points in the storyline, influencing interactions across various groups. This view helps in understanding the social structure within the Marvel universe and identifying key hubs or clusters in the network.
  - Color Insight: The Viridis palette creates a visually accessible scale from low (purple) to high (yellow) connection density, making central figures stand out. The use of a perceptually uniform Viridis palette ensures clarity for colorblind users and maintains contrast for easy identification of influential network hubs.

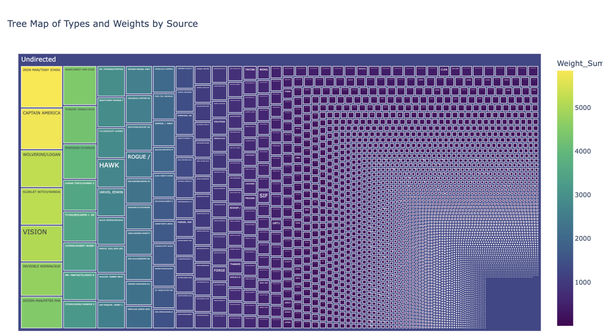

Fig. 48: TreeMap: Marvel Social Network Hierarchy based on degrees

3) Summary: Parallel Coordinate Plots and Treemaps are used to analyze student attrition trends in Australia and social connections within the Marvel universe. The PCPs reveal attrition patterns across demographics and qualifications, with interactive brushing allowing detailed exploration of specific groups. Treemaps provide hierarchical insights by organizing data by year and attribute, using a red-to-blue color scale to indicate risk levels—red for high attrition and blue for lower rates. A final treemap, using a Viridis palette, visualizes character connections in the Marvel network, highlighting key figures and clusters effectively.

VI. Visualizations

Following are the visualizations that are used and described in detail in the section above.

1) HeatMap
2) Quiver Plot
3) Streamline Plot
4) Contour Plots
5) ForceAtlas Layout
6) Fruchterman-Reingold Layout
7) OpenOrd Layout
8) TreeMap
9) Parallel Coordinate Plot

Also in each of the types wherever applicable, we have employed various marks and channels for making the visualizations more expressive for someone to get the maximum insights at the first glance.

VII. Member wise contributions

As the sole team member, I independently handled all tasks for the project.
All tasks were completed individually to ensure comprehensive coverage and cohesive results for the final project and report.

References

[1] https://www.climatologylab.org/gridmet.html
[2] https://gephi.org/datasets/hero-social-network.gephi
[3] https://data.gov.au/dataset/ds-dga-38e6bbd3-a071-482b-8742-7c64824308c1/details?q=Higher%20Education%20Attrition%20Rates
[4] https://numpy.org/
[5] https://scitools.org.uk/cartopy/docs/latest/
[6] https://github.com/pydata/xarray
[7] https://imageio.readthedocs.io/
[8] https://pillow.readthedocs.io/
[9] https://gephi.org/
[10] https://plotly.com/javascript/