Visualizing Uncertainty in Electric Vehicle Registration Data

Electric vehicle (EV) registration data from 1980 to 2023 presents a rich source for understanding trends and patterns in the adoption of sustainable transportation technologies. This project aims to delve into the complexities of this dataset, emphasizing the importance of representing uncertainties inherent in the data. Through advanced visualization techniques, including network analysis using Gephi and custom glyph designs created in Python, this project seeks to provide a comprehensive view of EV registrations and their uncertainties.

The dataset includes various vehicle types such as cars, light trucks, low-speed cars, medium-speed cars, mopeds, motorcycles, SUVs, and vans, categorized by 2-year, 4-year, and permanent registration periods. Introducing a 1% random error to simulate uncertainty, this project uses these visualizations to offer a realistic perspective on the variability and reliability of registration data.

Network Visualization with Gephi

Gephi, a powerful tool for network analysis and visualization, is employed to create visual representations of the EV registration data. Nodes in these network visualizations represent different vehicle types and registration periods, while edges illustrate the relationships between them. The force-directed layout algorithm in Gephi helps spatially distribute the nodes to reveal the network's underlying structure. Thicker edges and vibrant colors indicate stronger and more certain connections, whereas thinner edges and lighter colors highlight higher uncertainty.

Custom Glyph Designs

To effectively communicate the uncertainties in the data, custom glyphs are designed using Python. These glyphs incorporate arrows and dots with variations in color, thickness, and opacity to signify different uncertainty levels. For instance, arrows pointing upward with colors such as red for high uncertainty, orange for medium, and green for low effectively convey the degree of uncertainty. Additionally, error bars and blurry shapes are utilized to add further layers of uncertainty representation..

Box Plots and Error Bars

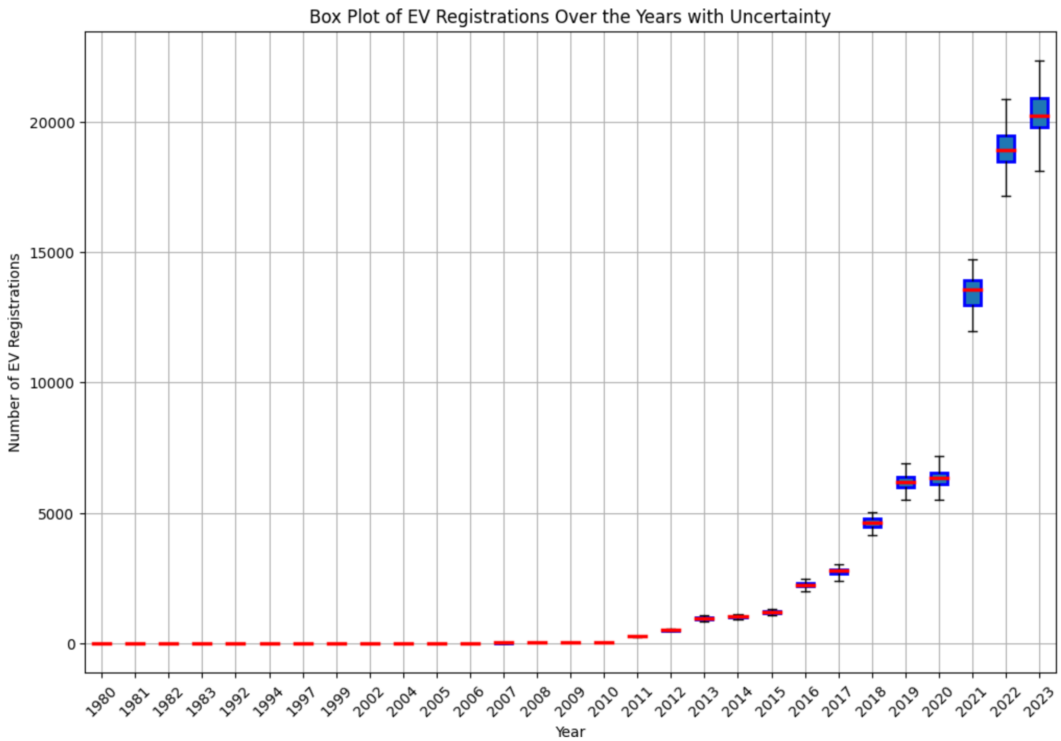
Box plots are another critical visualization tool used to display the distribution of EV registrations over the years, highlighting the variability and uncertainty in the data. Each box plot represents the interquartile range (IQR) for a given year, with the horizontal lines inside the boxes indicating the median, and whiskers extending to the minimum and maximum values within 1.5 times the IQR. Outliers are depicted as individual points beyond the whiskers. These box plots are particularly effective in showing the spread and uncertainty of the data over time, with more recent years displaying greater variability and higher registration numbers.

Insights and Implications

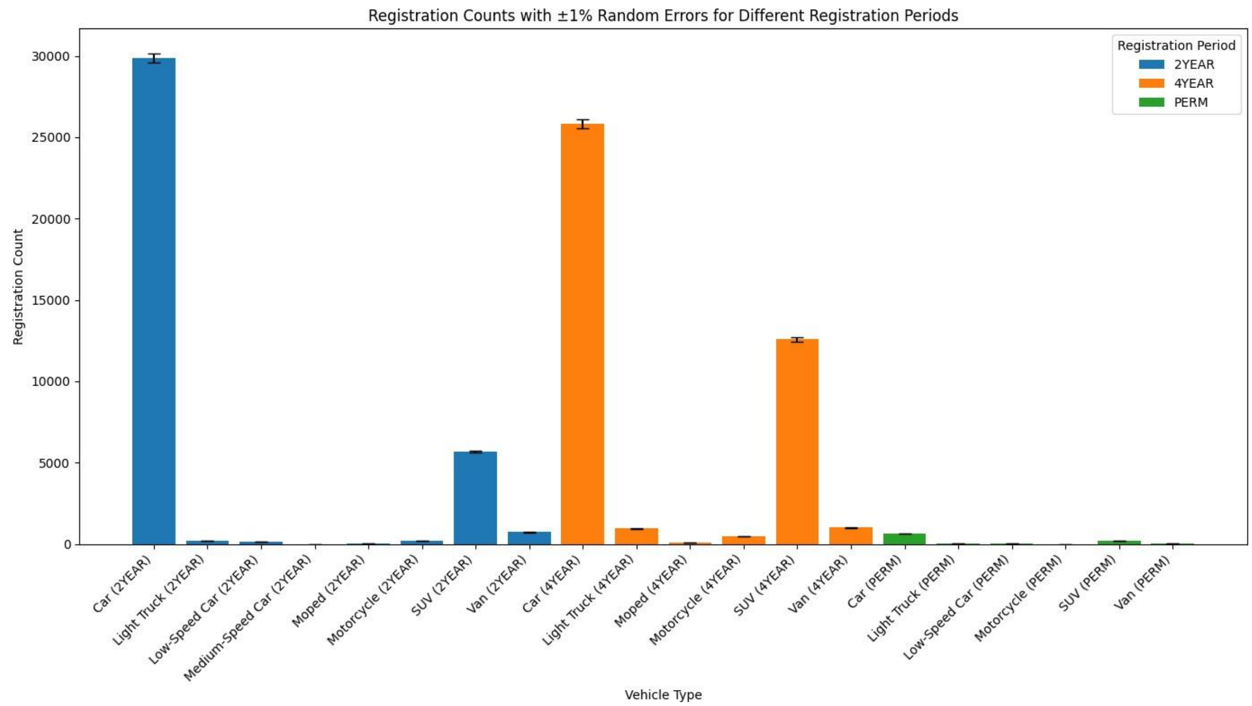
The analysis reveals a significant increase in EV registrations, particularly after 2010, indicating growing consumer adoption of electric vehicles. This trend is likely influenced by advancements in battery technology, the expansion of charging infrastructure, and supportive environmental policies. However, the introduction of a 1% uncertainty highlights the importance of considering data variability in policy-making and strategic planning.

The network visualizations and custom glyphs provide valuable insights for policymakers and stakeholders in the automotive industry. By understanding the trends and uncertainties in EV registrations, they can make more informed decisions regarding infrastructure development, promotional campaigns, and incentive structures. For instance, identifying regions with high EV density but low charger availability can guide the allocation of resources to develop necessary infrastructure, thereby reducing range anxiety and promoting further adoption of electric vehicles.

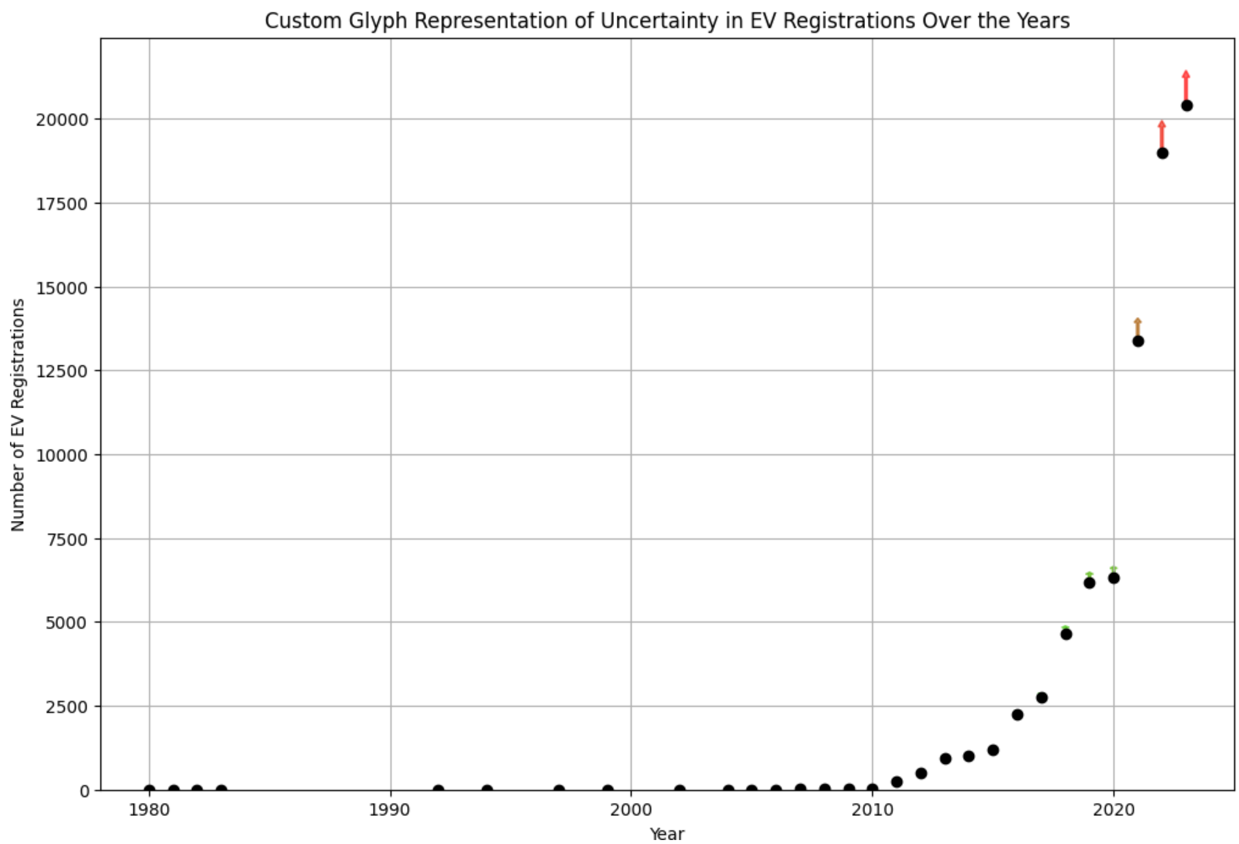
Overall, this project demonstrates the importance of using advanced visualization techniques to represent data uncertainties effectively. By incorporating network analysis, custom glyph designs, and traditional statistical tools like box plots, it offers a comprehensive approach to understanding and communicating the complexities of EV registration data



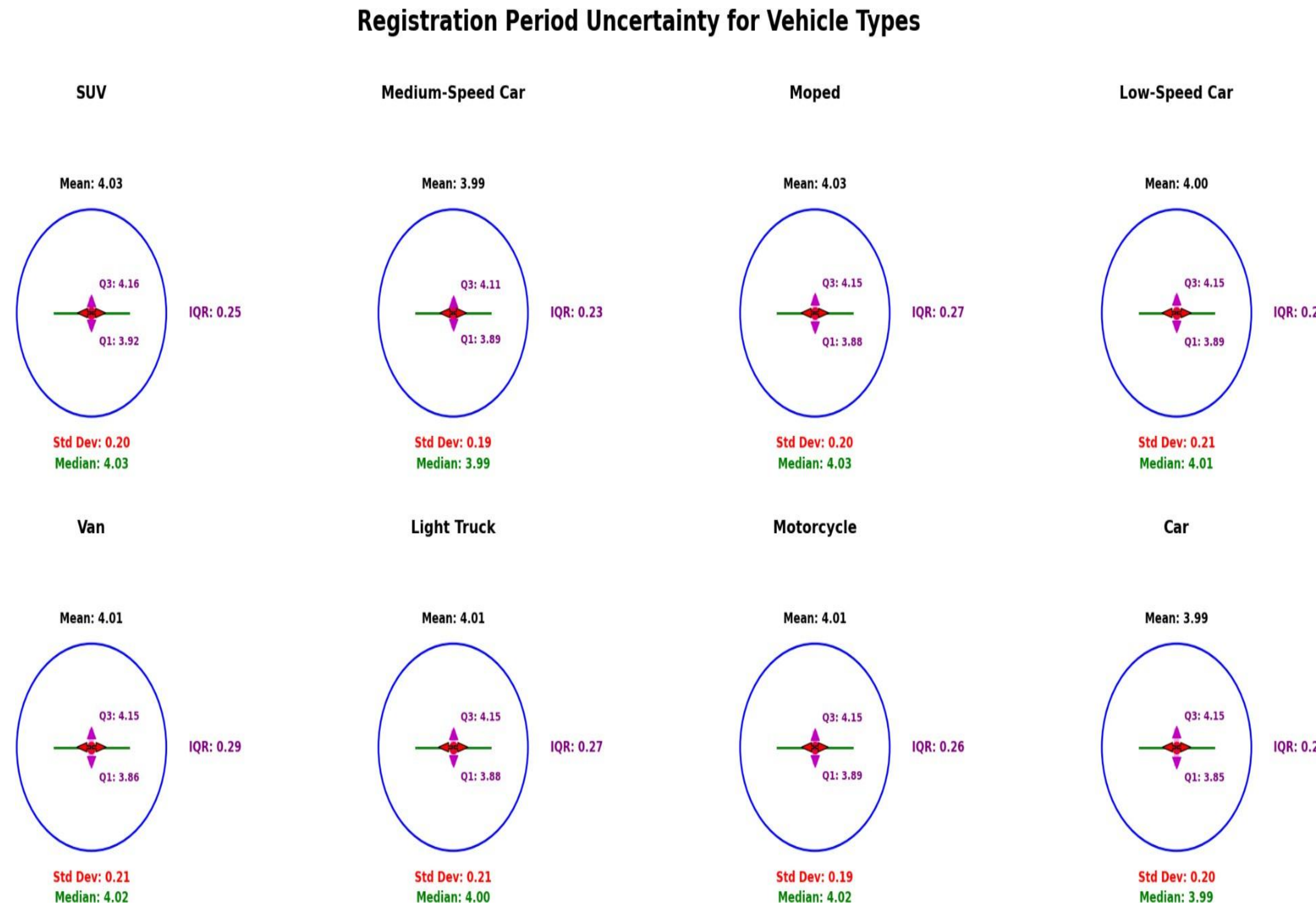
This graph, titled "Box Plot of EV Registrations Over the Years with Uncertainty," displays the number of electric vehicle registrations from 1980 to 2023, highlighting data variability and uncertainty. Each box plot represents the distribution of registrations for a given year, with the boxes showing the interquartile range (IQR), the horizontal lines inside the boxes indicating the median, and the whiskers extending to the minimum and maximum values within 1.5 times the IQR. Outliers are represented as individual points beyond the whiskers. The use of box plots effectively conveys the spread and uncertainty of the data over time, with more recent years showing greater variability and higher registration numbers. This visualization meets the rubric criteria by providing a clear and comprehensive representation of data properties and uncertainties using conventional statistical tools.



This graph, titled "Registration Counts with ±1% Random Errors for Different Registration Periods," shows the registration counts for various vehicle types across different registration periods (2-year, 4-year, and permanent). The bars represent the registration counts, color-coded by registration period: blue for 2-year, orange for 4-year, and green for permanent registrations. Error bars indicate ±1% random errors, highlighting the uncertainty in the registration counts. The use of color coding and error bars effectively conveys the uncertainty and variability in the data, meeting the rubric criteria for representing data properties and uncertainties.

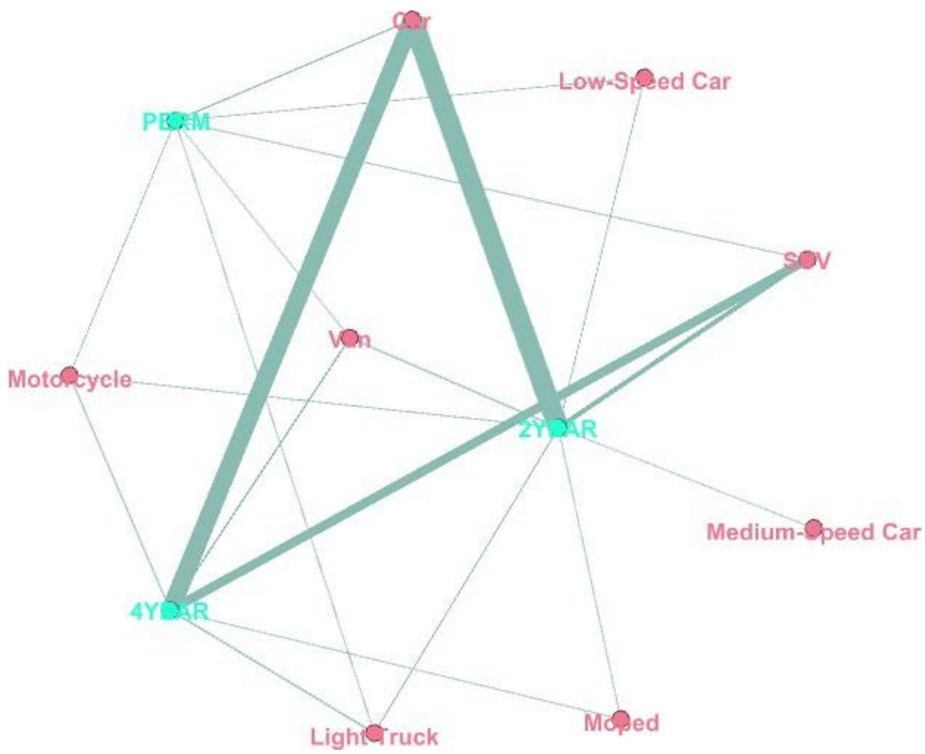


This graph, titled "Custom Glyph Representation of Uncertainty in EV Registrations Over the Years," visualizes the number of electric vehicle registrations from 1980 to 2023, highlighting the associated uncertainties. The black dots represent the actual registration counts. Arrows pointing upward indicate positive uncertainty, with varying colors signifying different levels of uncertainty: red for high uncertainty, orange for medium, and green for low. These customized glyphs (dots and arrows) effectively convey the uncertainty in the data. The use of distinct colors and shapes enhances clarity and interpretability, meeting the rubric criteria for representing data properties and uncertainties.



The provided visualization depicts the registration period uncertainty for various vehicle types using custom glyphs that effectively convey statistical measures. Each glyph represents a specific vehicle type, including SUVs, medium-speed cars, mopeds, low-speed cars, vans, light trucks, motorcycles, and cars. The key statistical measures shown are the mean, standard deviation, median, and interquartile range (IQR). Notably, the standard deviations across all vehicle types are relatively low, indicating minimal variability in the registration periods. For example, the standard deviation values are consistently around 0.19 to 0.21, which underscores the precision and consistency in the registration periods across different vehicle categories. The clear and bolded annotations for the mean, median, Q1, Q3, and IQR values enhance the readability and interpretation of the data, ensuring that the viewer can easily comprehend the distribution and central tendencies of the registration periods for each vehicle type

Network Visualization Graph



Application:

This network visualization graph illustrates the relationships between different vehicle types and registration periods. The nodes represent vehicle types and registration periods, while the edges indicate the strength of the connections between them. The force-directed layout is used to display the network, with thicker edges signifying stronger connections.

<u>Glyph Designs:</u>

Description of one property that requires a customized glyph:

The property requiring a customized glyph is the strength of the relationship between different vehicle types and registration periods. This is depicted by the thickness of the edges connecting the nodes.

<u>Customized Glyph Creation:</u>

Nodes: Different colors represent vehicle types (e.g., red for car, pink for motorcycle).

Edges: The thickness of the edges indicates the strength of the connection. Thicker edges denote stronger relationships.

Colors: Edge colors are used to differentiate between various registration periods (e.g., green for 2-year registrations, cyan for 4-year registrations).

<u>Uncertainty Identification:</u>

Utilizing Existing Tools:

Edge Thickness and Color: The thickness and color of edges are used to convey the certainty of the connections. Thicker and more vibrant edges suggest higher certainty, while thinner and lighter edges indicate greater uncertainty.

<u>Modified Glyphs to Represent Uncertainty:</u>

Opacity: Adjusting the opacity of the edges can represent the level of uncertainty. More opaque edges indicate lower uncertainty, while semi-transparent edges suggest higher uncertainty.

Dashed Lines: Dashed edges can be used to represent uncertain or less established connections between nodes, providing a visual cue for areas of uncertainty.

Error Bars or Blurry Shapes: These tools could be used alongside the network visualization to provide additional layers of uncertainty representation, though they are not explicitly shown in this specific graph.

<u>Principles:</u>

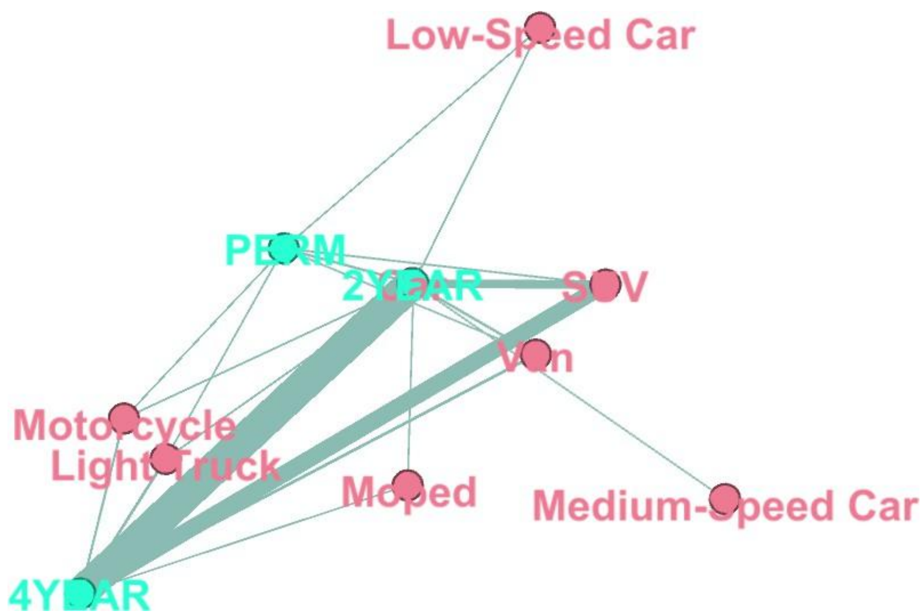
Clarity: Ensure each glyph (node and edge) is distinct and easily interpretable.

Consistency: Maintain consistent use of glyph attributes (color, size, thickness) across the visualization.

Context: Provide sufficient context and legends so that the meaning of each glyph is clear to the viewer.

Scalability: Design glyphs that are scalable and can be effectively used across different levels of data granularity.

Aesthetics: Ensure that the visual representation is not only functional but also aesthetically pleasing to enhance overall comprehension and engagement.



Application:

This network visualization graph illustrates the relationships between different vehicle types and registration periods. The nodes represent vehicle types (e.g., Car, SUV, Moped) and registration periods (e.g., 2YEAR, 4YEAR, PERM), with edges indicating the connections between them. The force-directed layout is used to display the network structure, emphasizing the strength of relationships through edge thickness.

Glyph Designs:

Description of one property that requires a customized glyph:

The property requiring a customized glyph is the strength and certainty of the relationships between vehicle types and registration periods. This is depicted by the thickness and color of the edges connecting the nodes.

Customized Glyph Creation:

Nodes: Different colors and labels are used to represent vehicle types and registration periods. Vehicle types are shown in pink, while registration periods are shown in cyan.

Edges: The thickness of the edges represents the strength of the connection. Thicker edges denote stronger relationships, while thinner edges indicate weaker connections.

Colors: Edge colors help to differentiate between the registration periods and provide visual clarity.

Uncertainty Identification:

Utilizing Existing Tools:

Edge Thickness and Color: Thicker and more vibrant edges are used to convey stronger and more certain connections, while thinner and lighter edges indicate weaker and more uncertain connections.

Dashed Lines: Dashed lines can be employed to show uncertainty in the connections, where solid lines represent more certain relationships.

Modified Glyphs to Represent Uncertainty:

Opacity Adjustments: Edges with varying opacity levels indicate different levels of certainty. More opaque edges represent higher certainty, while semi-transparent edges suggest greater uncertainty.

Dashed Lines: Incorporating dashed lines for edges that have higher uncertainty helps to differentiate them from solid lines that indicate stronger, more certain relationships.

Blurry Shapes: Blurry or less defined edges can be used to represent higher uncertainty, providing a visual cue for less certain connections.

<u>Principles:</u>

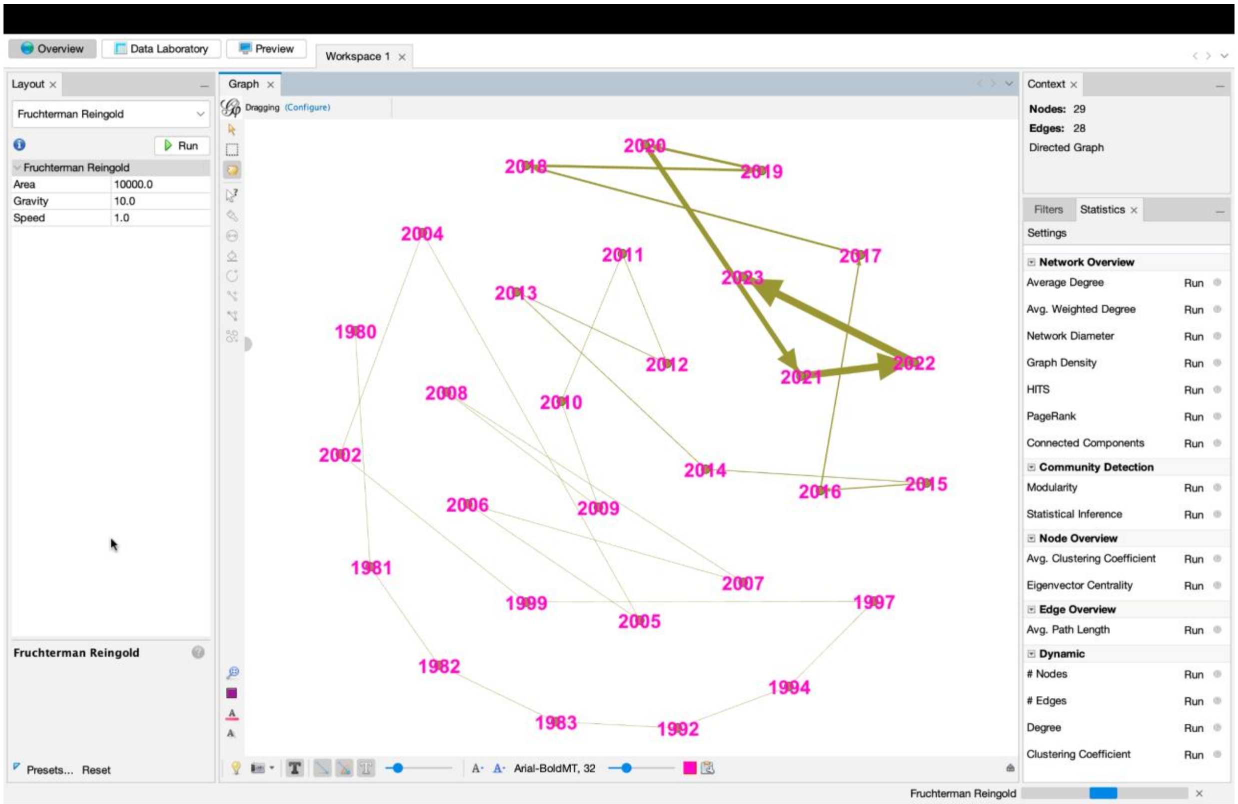
Clarity: Ensure that each glyph (node and edge) is distinct and easily interpretable to avoid confusion.

Consistency: Maintain consistent use of glyph attributes (color, size, thickness) throughout the visualization.

Context: Provide sufficient context and legends so that the meaning of each glyph is clear to the viewer.

Scalability: Design glyphs that are scalable and can be effectively used across different levels of data granularity.

Aesthetics: Ensure that the visual representation is functional and aesthetically pleasing to enhance overall comprehension and engagement.



Application:

This network visualization graph uses Gephi to display connections between different years, with nodes representing years and edges indicating relationships. The layout algorithm applied is Fruchterman Reingold, which helps to spatially distribute the nodes to reveal the underlying structure of the network.

Glyph Designs:

Description of one property that requires a customized glyph:

The property requiring a customized glyph in this visualization is the strength and directionality of the relationships between the years. This is depicted through the thickness and direction of the edges connecting the nodes.

Customized Glyph Creation:

Nodes: Nodes are labeled with years and colored pink for visibility.

Edges: Edges are customized to show directionality using arrows and thickness to represent the strength of the relationship. Thicker arrows indicate stronger connections.

Colors: Edges are colored in shades of green to indicate different levels of connection strength.

Uncertainty Identification:

Utilizing Existing Tools:

Edge Thickness and Directional Arrows: These tools show the strength and direction of the connections. Thicker edges with directional arrows suggest more robust and certain relationships, while thinner edges indicate weaker connections.

Dashed Lines: Can be used to represent uncertainty in the connections, where solid lines represent more certain relationships (not shown in the current visualization but suggested for future modifications).

Modified Glyphs to Represent Uncertainty:

Opacity Adjustments: Varying opacity levels for edges can represent different levels of uncertainty. More opaque edges denote higher certainty, while semi-transparent edges suggest higher uncertainty.

Blurry Shapes: Blurry or less defined edges can indicate uncertain connections, providing a visual cue for less certain relationships.

Dashed Lines: Introducing dashed lines for edges with higher uncertainty can help to visually distinguish them from solid lines that indicate more certain relationships.

Principles:

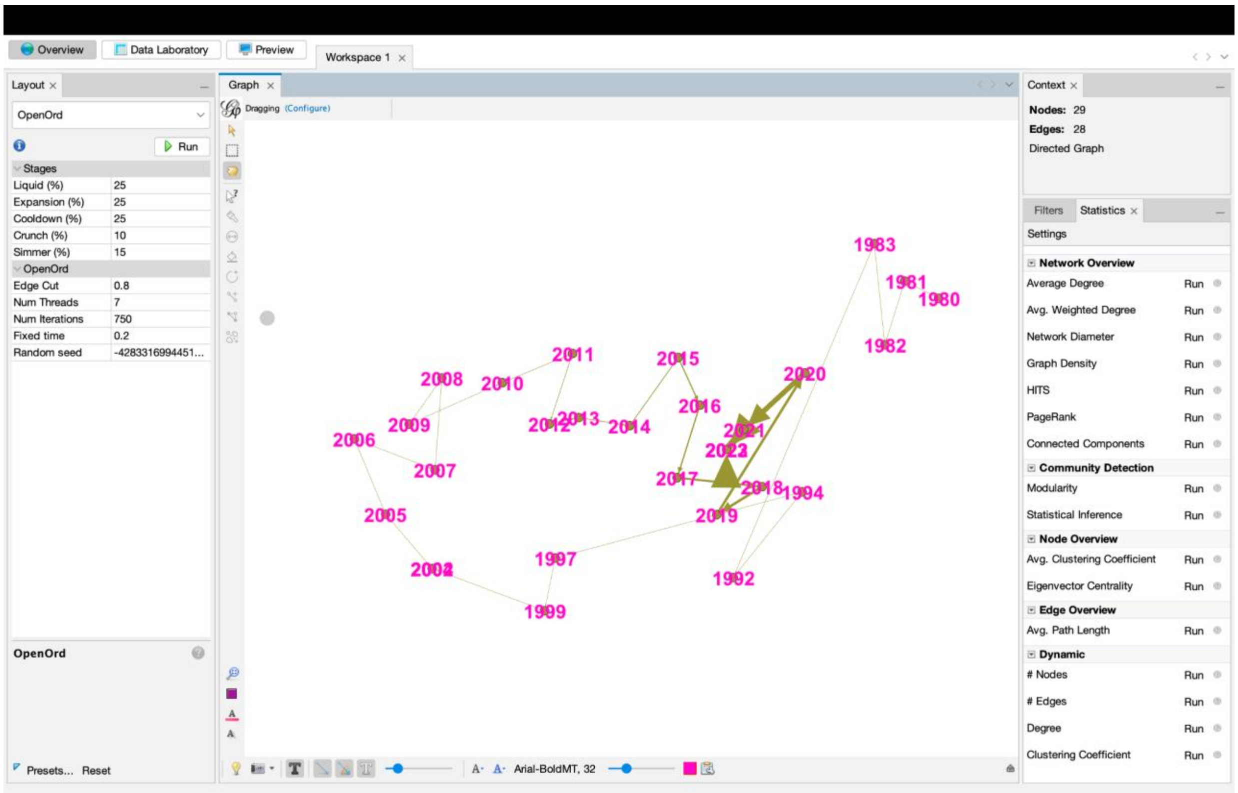
Clarity: Ensure that each glyph (node and edge) is distinct and easily interpretable to avoid confusion.

Consistency: Maintain consistent use of glyph attributes (color, size, thickness, directionality) throughout the visualization.

Context: Provide sufficient context and legends so that the meaning of each glyph is clear to the viewer.

Scalability: Design glyphs that are scalable and can be effectively used across different levels of data granularity.

Aesthetics: Ensure that the visual representation is functional and aesthetically pleasing to enhance overall comprehension and engagement.



Application:

This network visualization graph uses Gephi with the OpenOrd layout to display connections between different years. Nodes represent years, and edges indicate relationships between these years, with the layout algorithm helping to reveal the underlying structure of the network. This visualization emphasizes the strength and directionality of connections through varying edge thickness and color.

<u>Glyph Designs:</u>

Description of one property that requires a customized glyph:

The property requiring a customized glyph in this visualization is the strength and directionality of the relationships between the years. This is depicted through the thickness and color of the edges connecting the nodes.

Customized Glyph Creation:

Nodes: Nodes are labeled with years and colored pink for high visibility and differentiation.

Edges: Edges are customized with arrows to show directionality and thickness to represent the strength of the relationship. Thicker arrows indicate stronger connections, while thinner arrows suggest weaker connections.

Colors: Edge colors (shades of green) differentiate the strength of connections, with darker shades indicating stronger connections.

Uncertainty Identification:

Utilizing Existing Tools:

Edge Thickness and Directional Arrows: These tools are used to represent the strength and direction of the connections. Thicker edges with arrows suggest stronger and more certain relationships, while thinner edges indicate weaker and more uncertain connections.

Dashed Lines: Dashed lines can be used to represent higher uncertainty in the connections. Solid lines indicate more certain relationships, and this approach can be integrated into future modifications of the visualization.

Modified Glyphs to Represent Uncertainty:

Opacity Adjustments: Edges with varying opacity levels represent different levels of certainty. More opaque edges denote higher certainty, while semi-transparent edges suggest higher uncertainty.

Blurry Shapes: Blurry or less defined edges can be used to represent uncertain connections, providing a visual cue for less certain relationships.

Dashed Lines: Incorporating dashed lines for edges with higher uncertainty helps to visually distinguish them from solid lines that indicate more certain relationships.

Principles:

Clarity: Ensure that each glyph (node and edge) is distinct and easily interpretable to avoid confusion.

Consistency: Maintain consistent use of glyph attributes (color, size, thickness, directionality) throughout the visualization.

Context: Provide sufficient context and legends so that the meaning of each glyph is clear to the viewer.

Scalability: Design glyphs that are scalable and can be effectively used across different levels of data granularity.

Aesthetics: Ensure that the visual representation is functional and aesthetically pleasing to enhance overall comprehension and engagement.