Sankey Chart using Python

Mrs. Ansu Elsa Koshy

Karthick B Sarvesh J.A Yugapathy R.K

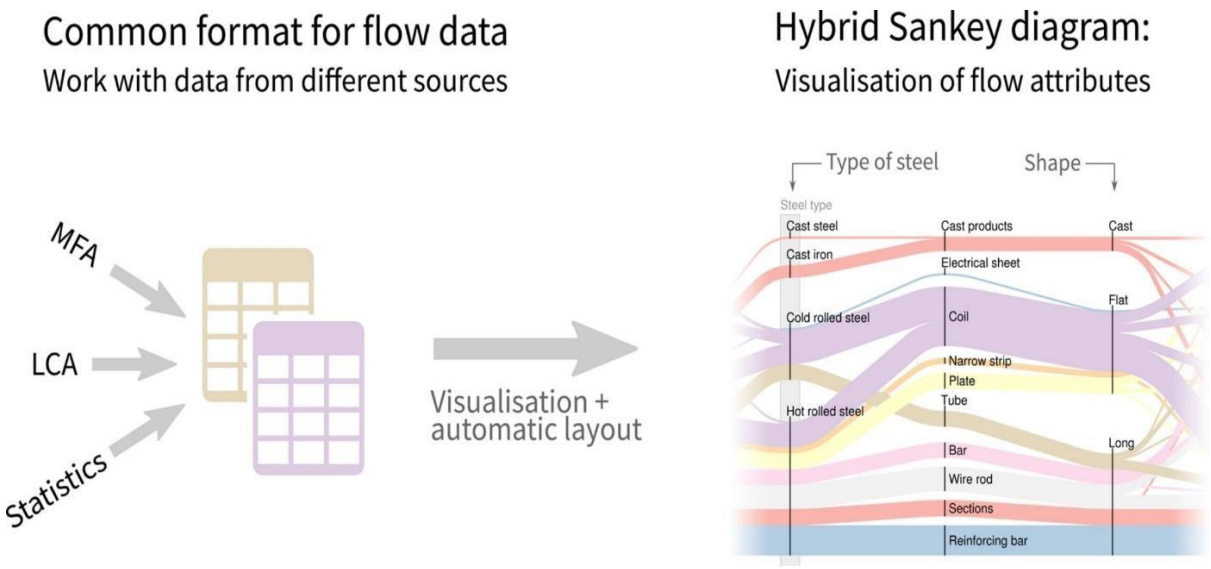
Bachelor Of Technology-2nd Year Department Of Artificial Intelligence And Data Science

Shakthi Institute Of Engineering And Technology (Autonomous) Coimbatore 641062

ABSTRACT

Sankey diagrams are a useful tool for mapping out manufacturing processes, helping to identify losses and inefficiencies, visualising material and energy movements, and providing a feeling of scale throughout a system. It can be necessary to use new visualisation techniques when available data and models get more intricate and sophisticated. For instance, it is not sufficient to look at only the flows of steel when searching for chances to reduce steel trash through supply chain integration; the alloy, thickness, coating, and forming history of the metal can all be crucial factors. In order to create a new kind of "hybrid" Sankey diagram that is more effective in visualising these many features of flows, this article mixes data-visualization techniques with the standard Sankey diagram.A Sankey diagram can be created in a number of ways, and different approaches are suitable in various contexts. In order to make this easier, an open-source Python implementation and a methodical approach for creating various hybrid Sankey diagrams from a dataset are provided. This method can be applied to construct Sankey diagrams from many data sources, including material flow analysis, life-cycle inventories, or directly measured data, by defining a standard data format for flow data. A number of graphic examples are used to demonstrate the technique, which is then applied to a genuine database of global steel flows.Sankey charts offer a dynamic and understandable depiction of intricate systems, flows, or processes. They are frequently recognised as an effective visualisation technique. These diagrams gracefully illustrate how energy, resources, and information are distributed inside a system, providing a comprehensive perspective that makes in-depth examination easier. Sankey charts are recognisable by their unique flow diagrams, which use different widths of connected routes to represent the amount or size of the objects being tracked. The intricate layout of these diagrams enables interested parties to spot trends, pinpoint obstructions, and maximise productivity, rendering them an indispensable resource for decision-makers in domains spanning from energy administration and manufacturing procedures to monetary dealings and data flow operations. Sankey charts essentially act as a visual bridge, shedding light on the complex relationships within complex systems.

Graphical abstract



CHAPTER 1

INTRODUCTION

OVERVIEW

A complex data visualisation technique called a "sankey chart" is used to show how information, energy, or resources move through complex systems. These maps use a distinct visual language in which the number or size of components being tracked is represented by a network of routes with different widths. Sankey charts are widely used in a variety of industries, including energy management, finance, industrial processes, and information workflows. They provide a thorough and understandable depiction of complex interactions. Their capacity to uncover trends, pinpoint bottlenecks, and maximise efficiencies renders them a priceless tool for decision-makers pursuing a more profound comprehension of intricate systems.

The ability of Sankey charts to give stakeholders a comprehensive understanding of interrelated processes is one of their main advantages. These charts help viewers understand the dynamics at work in a system by graphically representing the paths and relative sizes of various components. By exposing areas of concentration, detecting inefficiencies, and assisting in the optimisation of resource allocation, this holistic approach makes strategic decision-making easier. Sankey charts therefore offer as a link between data and practical insights, enabling businesses to make wise decisions and boost productivity.

Furthermore, Sankey charts can be used in a wide range of businesses, making them a useful tool for data analysis. Sankey charts are an adaptable and perceptive way to represent complicated interactions, whether you're tracking financial transactions in a corporation, illustrating the flow of energy in a power plant, or visualising information workflows in a digital platform. The versatility of Sankey charts and their capacity to simplify complex information into an understandable manner highlight their significance in supporting decision-makers in a range of industries as they strive for resource and process optimisation.

GENERAL INFORMATION

Sankey charts are a specialized type of flow diagram that visualizes the flow of resources, energy, or information within a system. Named after Captain Matthew Henry Phineas Riall Sankey, who first introduced this graphical technique in the late 19th century, Sankey charts have evolved into a powerful tool for presenting complex data in a clear and concise manner. The fundamental concept revolves around the representation of flows as interconnected pathways, with the width of each path corresponding to the quantity ormagnitude of the flow. This visual depiction enables users to grasp the distribution and movement of elements within a system, making Sankey charts particularly useful in various fields such as engineering, economics, and environmental sciences.

Because of their adaptability and clarity, simplex charts are used extensively across a wide range of disciplines and industries. They are frequently used to improve data communication and provide a deeper comprehension of complicated systems in dashboards, reports, and presentations. Sankey charts continue to be essential tools for converting complex numbers into readable, actionable information as more and more businesses realise the value of making decisions based on data.

OBJECTIVE

The primary objectives of employing Sankey charts are rooted in their capacity to provide clear and insightful visual representations of complex systems, facilitating enhanced understanding and decision-making. Firstly, Sankey charts aim to visually convey the distribution and flow of resources, energy, or information within a system. By employing interconnected paths with varying widths, these charts offer a nuanced depiction of quantities, enabling stakeholders to identify patterns and relationships. This visual clarity serves as a crucial tool for comprehending the intricate dynamics of a system. Secondly, Sankey charts aim to pinpoint inefficiencies, bottlenecks, or areas of optimization within a given process or system. The visual hierarchy and proportional representation of flows allow users to quickly identify areas where resources are underutilized or where there may be constraints. This objective aligns with the overarching goal of improving efficiency and resource allocation, making Sankey charts a valuable tool for process optimization.

Lastly, the objective of Sankey charts is to enhance communication and decision-making by presenting complex data in an accessible and intuitive format. The visual impact of Sankey charts aids in conveying key insights to diverse audiences, facilitating better- informed decisions. Whether used in environmental sciences, engineering projects, or financial analyses, the ultimate aim is to empower stakeholders with a comprehensive and visually compelling understanding of complex systems, contributing to more effective problem-solving and strategic planning.

CHAPTER 2

LITERATURE REVIEW

Introduction: The Power of Flow Visualization

Sankey diagrams have become an indispensable tool for visualizing flows and relationships between entities, owing to their unique ability to simplify complex data. This section explores the underlying principles that make Sankey diagrams a preferred choice in various fields. By offering a balance of simplicity and clarity, these diagrams make intricate data patterns accessible to a broad audience. The discussion delves into the visual storytelling prowess of Sankey charts, emphasizing their effectiveness in conveying complex information through intuitive flow visualization. Sankey charts have become a pivotal tool in visual storytelling and data representation. This section introduces the significance of Sankey charts in conveying complex relationships and flows within a system. Sankey diagrams, with their distinct flow paths and proportional widths, offer a visually intuitive means to communicate intricate data patterns, making them a popular choice in various fields.

Evolution of Sankey Charts: A Historical Journey

The evolution of Sankey charts is explored, tracing their historical development and adaptation. Originally introduced by Captain Matthew Henry Phineas Riall Sankey in the late 19th century, these diagrams have undergone significant refinement. Over time, advancements in technology and data visualization techniques have contributed to the versatility and widespread use of Sankey charts, making them a fundamental component of modern visual analytics. The historical evolution of Sankey charts is a fascinating journey that unfolds from their inception by Captain M.H.P.R. Sankey in the late 19th century. This subsection takes readers on a chronological exploration, tracing the transformation of these diagrams from rudimentary visualizations to the sophisticated tools we encounter today. Technological advancements and innovations in data visualization techniques are examined as key factors shaping the versatility and widespread use of Sankey charts.

Applications of Sankey Diagrams: A Spectrum of Possibilities

This subsection delves into the diverse applications of Sankey diagrams in the realm of data visualization. From energy management and industrial processes to financial transactions and information workflows, Sankey charts prove to be versatile tools. The discussion highlights how these charts aid decision-makers in understanding complex systems, optimizing processes, and conveying insights in a visually compelling manner. This section illuminates the diverse applications of Sankey diagrams, showcasing their versatility across various domains. Energy management, industrial processes, financial transactions, and information workflows emerge as key areas where Sankey diagrams excel. Each application is briefly explored, emphasizing how Sankey diagrams contribute to visualizing and optimizing complex systems.

CHAPTER 3

RESEARCH METHODOLIGIES

Introduction:

The introduction to a research methodology is a comprehensive narrative that serves as the roadmap for the entire study. It goes beyond merely stating the problem by contextualizing it within the broader field of research. This section begins by outlining the background and historical context of the issue at hand, providing readers with a clear understanding of its evolution and relevance. The researcher then articulates the specific research problem, elucidates its significance, and outlines the objectives that the study aims to achieve. The introduction also incorporates a concise review of relevant literature, showcasing the existing body of knowledge and identifying any gaps or contradictions that the current research seeks to address. Furthermore, the theoretical framework guiding the study is introduced, offering readers insights into the conceptual underpinnings shaping the research approach. A well-constructed introduction not only captivates the reader's attention but also provides a solid foundation for the subsequent sections of the research methodology.

Data Collection:

The data collection phase is a pivotal element in any research methodology, demanding meticulous planning and execution. This section delves into the intricacies of the data collection process, detailing the chosen methodology, sampling techniques, and tools employed. If the study involves primary data, researchers discuss how participants were selected, informed consent obtained, and ethical considerations addressed. In the case of secondary data, the sources, reliability, and relevance of the data are thoroughly examined. The rationale behind opting for a specific data collection method is expounded, addressing any potential biases and limitations. Researchers also discuss the challenges faced during data collection and the steps taken to mitigate them, ensuring the robustness and

reliability of the gathered information. A transparent account of the data collection process enhances the study's credibility, allowing other researchers to replicate or build upon the findings.

Sankey Chart Design: The Sankey chart design section focuses on the visual representation of data, particularly the intricate relationships between variables. Researchers discuss the rationale behind choosing a Sankey chart over other visualization methods, emphasizing its ability to convey complex data flows in a clear and intuitive manner. The choice of software or tools for creating the Sankey chart is justified, and the parameters influencing its design are outlined. The section provides a step-by- step guide to the creation process, ensuring transparency and enabling other researchers to replicate the visualizations. The visual appeal of the Sankey chart is considered alongside its ability to enhance data interpretation, making it an effective tool for presenting interconnected variables. The section also discusses potential challenges in Sankey chart design and strategies employed to address them, ensuring the accuracy and reliability of the visual representation. Ultimately, a well-designed Sankey chart not only enhances the research's visual appeal but also serves as a powerful tool for conveying complex data relationships to diverse audiences. In conclusion, each of these components plays a crucial role in shaping a robust research methodology, from establishing the foundation and gathering data to effectively visualizing and communicating the findings. A meticulous and transparent approach in these sections ensures the credibility, replicability, and overall success of the research endeavor.

Sankey Chart Evaluation

The evaluation of Sankey charts is an essential step in assessing their quality and utility. This subsection describes the criteria and methods used to evaluate the Sankey diagrams created in the research. It explains how the Sankey charts were compared and contrasted with other visualization techniques, as well as how the feedback from potential users and experts was collected and analyzed. The evaluation aims to measure the effectiveness of Sankey charts in terms of accuracy, clarity, aesthetics, and engagement.

CHAPTER 4

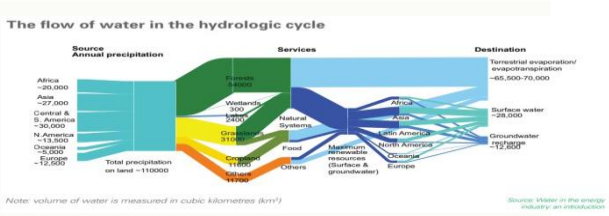
SYSTEM REQUIRMENTS

Web Browsers:
Sankey chart tools and applications are often web-based, necessitating compatibility with popular web browsers such as Google Chrome, Mozilla Firefox, Safari, or Microsoft Edge. Users can leverage these browsers to access and interact with Sankey charts seamlessly. Compatibility across different browsers ensures widespread accessibility and usability, facilitating a consistent experience for users across various platforms. Software **Dependencies:**
To create and render Sankey charts, users typically rely on software or programming libraries that provide the necessary functionalities. Common options include D3.js, Plotly, Highcharts, and others. Depending on the chosen tool, additional software dependencies may be required, such as Node.js or Python. Users should ensure that the selected software stack aligns with their technical stack and integrates seamlessly into their development environment.

Web Development Tools:
For users engaging in web-based Sankey chart development, familiarity with web development technologies is beneficial. Knowledge of HTML, CSS, and JavaScript is often required to customize and embed Sankey charts into web applications. This understanding of web development tools enhances the user's ability to integrate Sankey charts seamlessly into their web- based projects and applications. Data Format: Sankey charts typically require input data in specific formats, such as matrices or hierarchical structures. The chosen library or tool will dictate the exact format, and users must format their data accordingly. Understanding the required data format is crucial for accurately representing the flow of information or resources in the resulting Sankey chart. Graphics Acceleration (Optional): While not mandatory, graphics acceleration, often facilitated by a dedicated graphics processing unit (GPU) with

WebGL support, can enhance the rendering performance of Sankey charts. This is particularly relevant for handling large datasets or intricate visualizations, where improved graphics processing capabilities contribute to smoother interactions and a more responsive user experience.

RESULT:



CHAPTER 8
REFERENCE

Alemasoom, H., Samavati, F., Brosz, J., Layzell, D., 2015. EnergyViz: an interactive system for visualization of energy systems. Vis. Comput. 32, 403–413. http://dx.doi. org/10.1007/s00371- 015-1186-8. Allwood, J., Ralph, D., Richards, K., Fenner, R., Linden, P., Dennis, J., Gilligan, C., Pyle, J., Kopec, G., Bajželj, B., Curmi, E., Qin, Y., Lupton, R., 2017. Foreseer – integrated resource analysis. https://www.foreseer.group.cam.ac.uk. Bajželj, B., Allwood, J.M., Cullen, J.M., 2013. Designing climate change mitigation plans that add up. Environ. Sci. Technol. 47, 8062–8069. http://dx.doi.org/10.1021/ es400399h. Brunner, P.H., Rechberger, H., 2003. Practical Handbook of Material Flow Analysis. CRC Press. Cabrera Serrenho, A., Allwood, J.M., 2016. Material stock demographics: cars in Great Britain. Environ. Sci. Technol. 50, 3002–3009. http://dx.doi.org/10.1021/acs.est. 5b05012. Chaudhuri, S., Dayal, U., 1997. An overview of data warehousing and OLAP technology. SIGMOD Rec. 26, 65–74. http://dx.doi.org/10.1145/248603.248616. Chen, C., Yan, X., Zhu, F., Han, J., Yu, P.S., 2009. Graph OLAP: a multi-dimensional framework for graph data analysis. Knowl. Inf. Syst. 21, 41–
63. http://dx.doi.org/10. 1007/s10115-009-0228-9. Cullen, J.M., Allwood, J.M., 2010. The efficient use of energy: tracing the global flow of energy from fuel to service. Energy Policy 38, 75–81. http://dx.doi.org/10.1016/j. enpol.2009.08.054. Cullen, J.M., Allwood, J.M., 2013. Mapping the global flow of aluminum: from liquid aluminum to end-use goods. Environ. Sci. Technol. 47, 3057–3064. http://dx.doi. org/10.1021/es304256s. Cullen, J.M., Allwood, J.M., Bambach, M.D., 2012. Mapping the global flow of steel: from steelmaking to end-use goods. Environ. Sci. Technol. 46, 13048–13055. http://dx. doi.org/10.1021/es302433p. Curmi, E., Fenner, R., Richards, K., Allwood, J.M., Bajželj, B., Kopec, G.M., 2013. Visualising a stochastic model of Californian water resources using Sankey diagrams. Water Resour. Manage. 27, 3035–3050. http://dx.doi.org/10.1007/s11269-013- 0331-2. Density Design, 2010. Fineo.