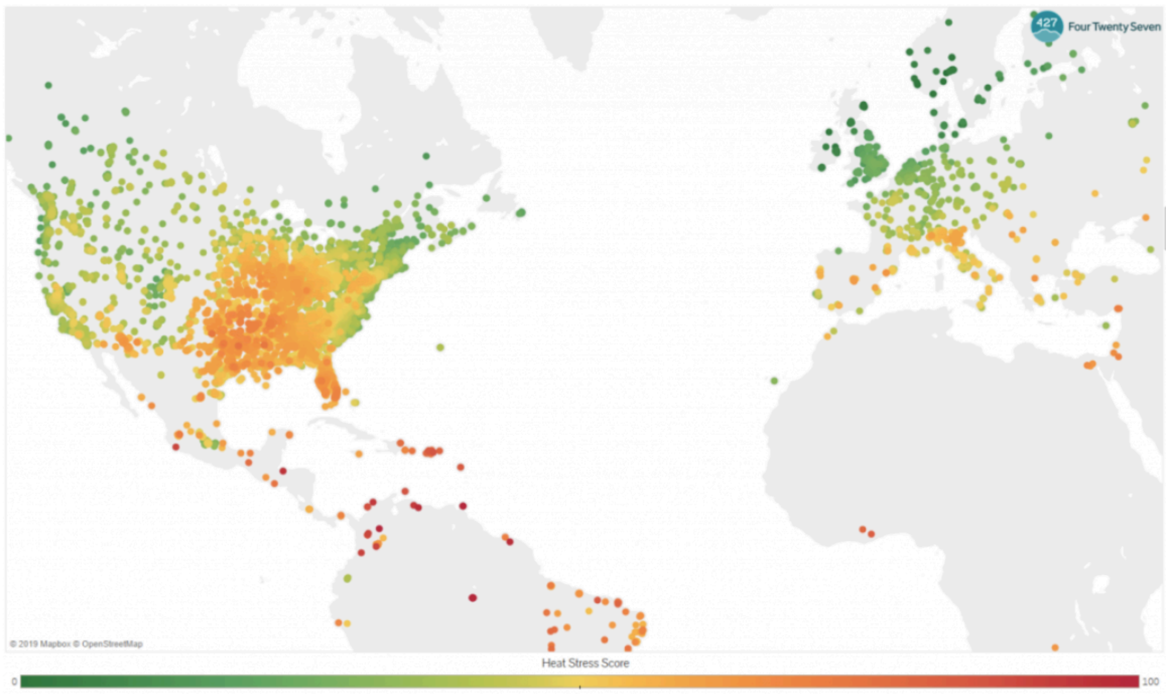
INFOSCI 301 – Data Visualization and Information Aesthetics
Prof. Luyao Zhang
Information Visualization Redesign Project
Yifei Yang
GitHub Repository: [YifeiCathyYang/Redesign-Project: INFO 301 Information Visualization Redesign Project]

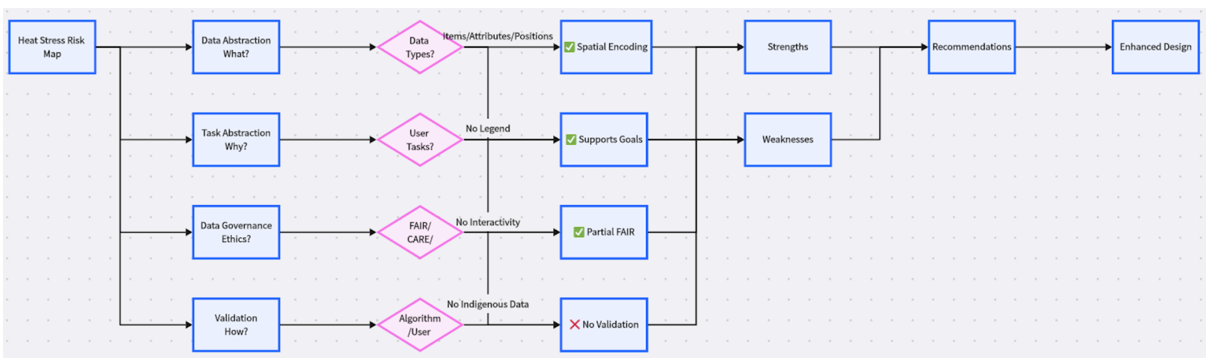
1. Theory – Critical Engagement with Visualization Methodologies

Figure 1. Heat Stress Risk: Global Exposure Assessment



Source: **Amazon Web Services**

Figure 2. Critique Process of the Heat Stress Risk Visualization



Source: Created by Boardmix

This reflection integrates fundamental visualization theories to evaluate both the strengths and weaknesses of the climate risk map. By applying key concepts such as data and task abstraction, encoding principles, and ethical data governance, this analysis highlights how the visualization succeeds in some areas while falling short in others.

One of the primary strengths of the map lies in its effective data abstraction. The map significantly enhances interpretability by transforming complex climate risk data from tabular formats into spatial visualizations. Risk levels are encoded as color gradients, aligning with established frameworks for data types (items, attributes, positions) and dataset types (geometry, fields), as discussed in Chapter 2 of Munzner (2014). Furthermore, according to Munzner (2014, Chapter 2), ordered attributes such as color gradients are effective for representing ordinal data. This heat visualization employs ordered attributes, such as a gradient color scale, to represent quantitative risk levels, which adheres to best practices for encoding ordinal data. It also demonstrates strong alignment with task abstraction principles. By supporting high-level user objectives, such as discovering regions of elevated risk and presenting climate vulnerabilities to stakeholders, both of which are consistent with Munzner's task abstraction framework (Munzner, 2014).

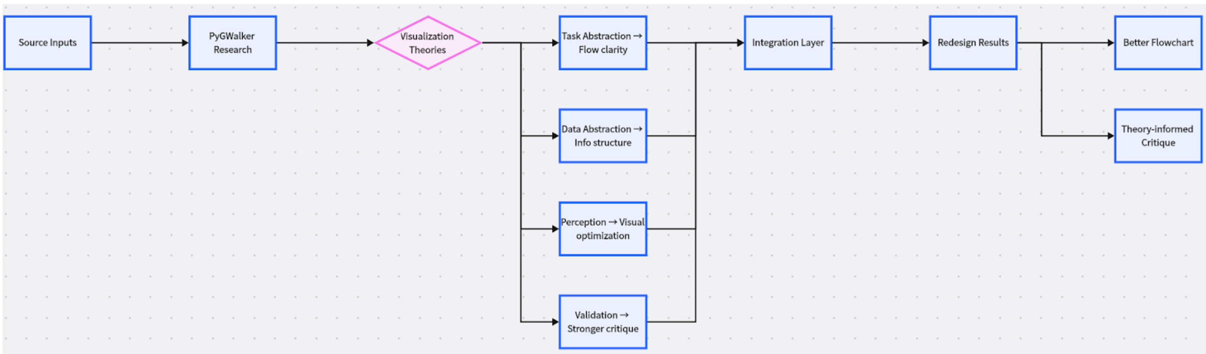
Despite its strengths, the visualization displays several weaknesses in visual encoding and data integrity. The absence of a legend violates fundamental visual encoding principles, hindering users' ability to interpret the mapping between color and risk level (Munzner, 2014). Additionally, the visualization misses the opportunity to employ hybrid encoding techniques, such as combining symbol size with color, to enrich data representation (Map Library, n.d.). The map also falls short in terms of data governance. Although the source of the data (Four Twenty Seven's database) is cited, essential aspects such as data accessibility and interoperability are not addressed, undermining adherence to the FAIR principles (Wilkinson et al., 2016). Another critical limitation lies in the lack of validation. As explained in Chapter 4 of Munzner (2014), idiom validation involves assessing the effectiveness of visual encoding and interaction choices in a visualization design. However, in this heat risk map, the methodology behind the risk score calculations, such as the weighting of heat metrics, is not transparent, which undermines the credibility and correctness of the visualization. Finally, dataset availability is another limitation, particularly in terms of static versus dynamic data. As discussed in Chapter 2 of Munzner (2014), dynamic data streams evolve over time and require visualizations that can accommodate temporal change. However, the static nature of this visualization fails to capture temporal dynamics, such as seasonal variations in climate risk, which diminishes its usefulness for ongoing decision-making.

2. Research – Literature-Inspired Analysis

Figure 3. Eight Facets of Research Process Funnels

Source: Created by Boardmix

Figure 4. Integration of Fundamental Visualization Theory and Research



Source: Created by Boardmix

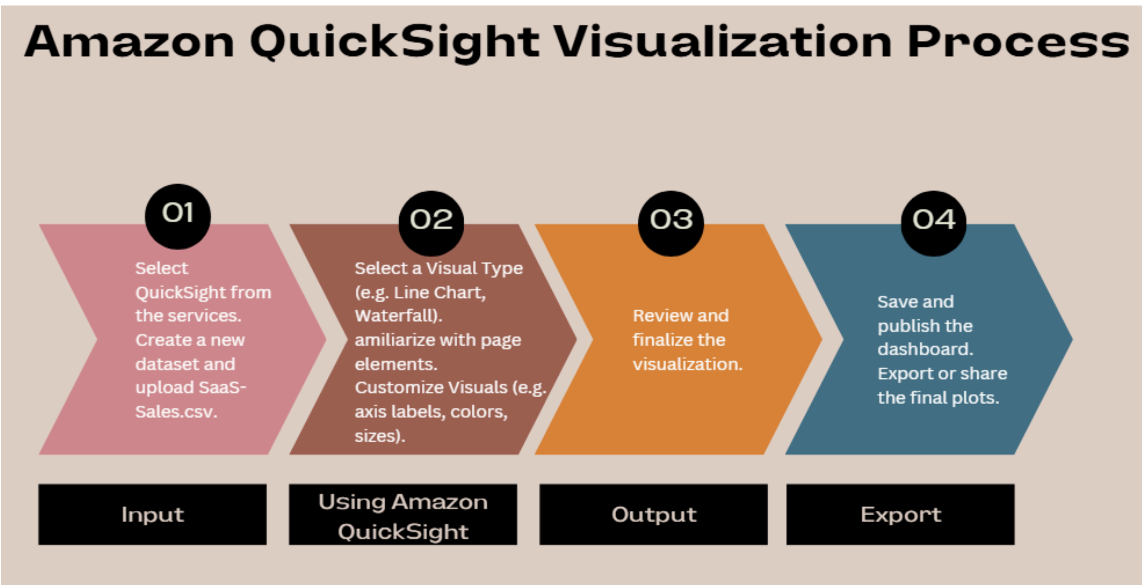
The flowchart illustrates the progression of the PyGWalker paper, beginning with its background and motivation, which aim to address the disconnect between programmatic and visual analysis in computational notebooks. The authors identify a gap that limits accessibility and speed in exploratory data analysis (EDA), especially for users without advanced coding skills. The research question centers on bridging this divide by developing a tool that enhances interactivity and accessibility.

The solution proposed is PyGWalker, a seamless bridge between Python code and interactive visualization through a GUI powered by Graphic Walker. The methodology section outlines its architecture, highlighting Graphic-Link and Compute-Link, which tightly couple visualization with code execution. These elements enhance perceptual optimization by aligning interface design with human visual cognition principles, such as preattentive processing and graphical encoding.

The results demonstrate PyGWalker's effectiveness and popularity through rapid adoption metrics, case studies, and user feedback. Intellectual merit lies in its theoretical contribution to human-in-the-loop analysis, while the practical impact spans both novice and expert users across data science domains. Compared to existing EDA tools like Voyager (Wongsuphasawat et al., 2016) and Lux (Lee et al., 2021), which offer visualization recommendations, PyGWalker stands out for its tight integration into the notebook ecosystem, streamlining the visualization process without requiring context switching. Drawing from the authors' emphasis on aligning interface elements with human visual cognition, the following redesign project prioritized perceptual optimization by employing consistent layout structures, spatial grouping, and intuitive color schemes to enhance preattentive processing and minimize cognitive load. These design choices reflect principles discussed in the paper's methodology, where PyGWalker leverages perceptually effective encodings to support rapid pattern detection (Wongsuphasawat et al., 2016). Furthermore, inspired by the paper's emphasis on accessibility for users with limited programming expertise, the redesign was enhanced through improved labeling strategies and the use of visual affordances to support intuitive interaction (Wongsuphasawat et al., 2016). These improvements aim to make the visualization more inclusive, usable, and aligned with the needs of both novice and expert audiences

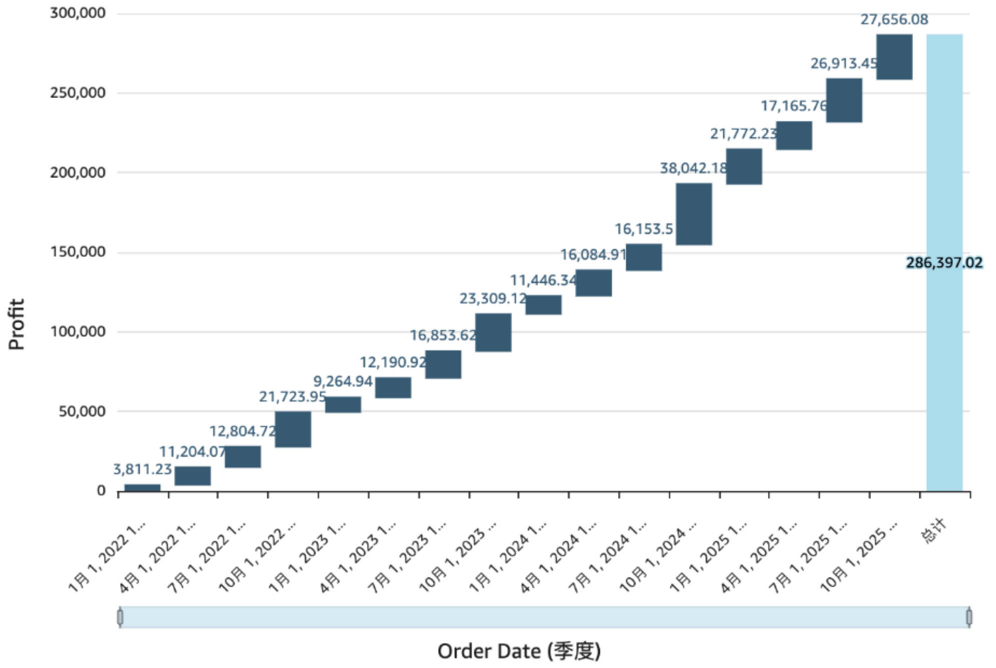
3.1 Practice – Amazon QuickSight Workshop

Figure 5. Workflow of the Visualization Process by Amazon QuickSight



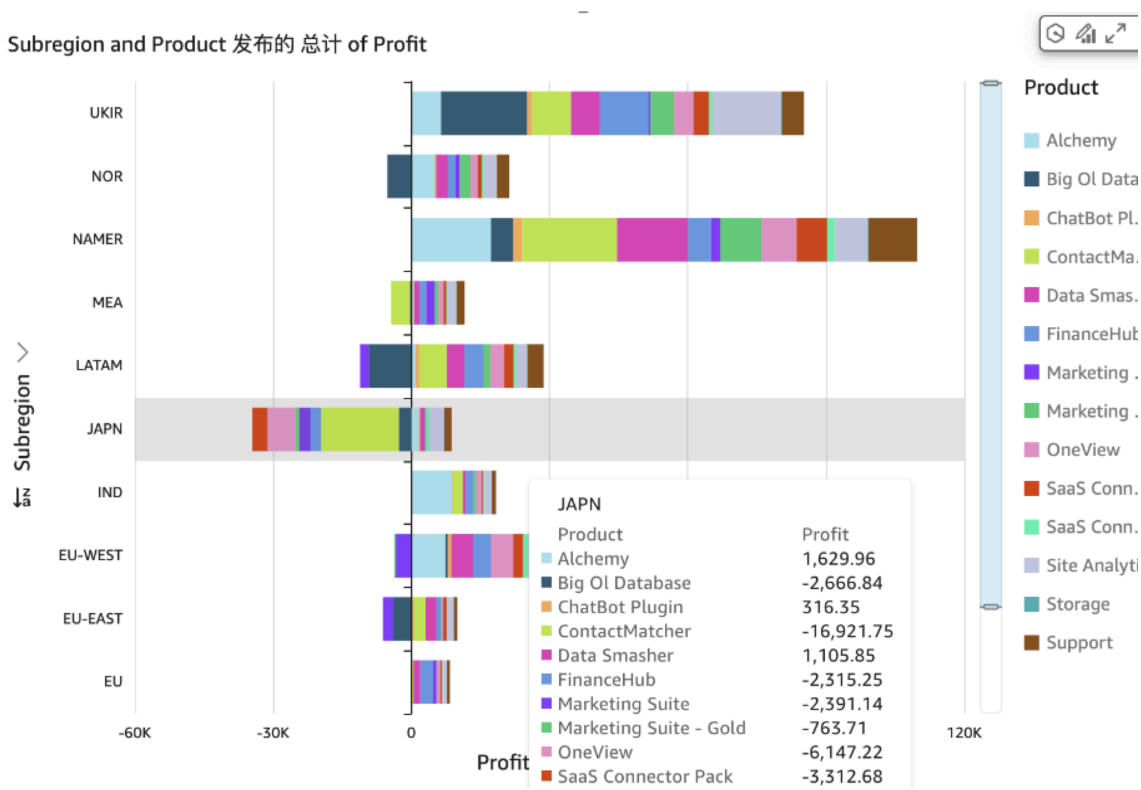
Source: Created by Canva

Figure 6. Waterfall of Quarterly Profit Growth Trend



Source: Created by Amazon QuickSight *(Explanation: The waterfall chart illustrates a steady increase in cumulative profit over time, with quarterly gains contributing to a total profit of 286,397.02. The consistent positive increments indicate sustained profitability, with some quarters showing larger profit jumps, such as Q4 2024 and Q1 2025.)*

Figure 7. Profit Distribution by Subregion and Product



Source: Created by Amazon QuickSight *(Explanation:* The distribution chart displays profit distribution by subregion and product, showing that different products contribute variably to the overall profit in each region. Notably, Japan (JAPN) has significant losses from "ContactMatcher" and "OneView," while some products like "Alchemy" contribute positively.)

The following sections offer a comprehensive overview of the project's goal, the tool utilized, the workflow guiding the analysis, the data insights, and the aesthetic and ethical reflection.

Goal:

This visualization project aims to present profit trends and distribution patterns to facilitate a clearer understanding of quarterly growth and regional/product-level profitability. Two core charts were developed: a waterfall chart to illustrate quarterly contributions to overall profit growth and a stacked bar chart to examine profit distribution across subregions and products. These visual tools are intended to generate insights that inform strategic business decisions and financial planning.

Tool Used:

The visualizations were created using Amazon QuickSight, a cloud-based business intelligence platform suitable for dynamic dashboards and interactive analytics.

Workflow:

- Data Preparation: Quarterly profit data were imported and cleaned to ensure format consistency and data completeness.
- Waterfall Chart Construction: A cumulative profit waterfall chart was designed to highlight how each quarter contributed to overall growth. Notable increases were observed in Q4 2024 and Q1 2025, indicating periods of accelerated profitability.
- Profit Distribution by Subregion and Product: A stacked bar chart was constructed to reveal how various subregions and product lines contributed to the overall profit. The visualization uncovered significant losses in Japan (JAPN), particularly from *ContactMatcher* and *OneView*, while products such as *Alchemy* demonstrated strong profitability in other regions.

Data Insight:

The waterfall chart confirmed a generally positive profit trajectory, punctuated by key quarters with substantial gains. The stacked bar chart brought attention to regional and product-specific discrepancies, revealing both high-performing areas and segments in need of strategic reassessment. These findings underscore the importance of reinforcing successful product lines and conducting further evaluation of underperforming regions.

Aesthetic Reflection:

- Color Differentiation: Contrasting color schemes were used to distinguish clearly between profit increases and losses, enhancing immediate pattern recognition.
- Labeling and Layout: Structured labeling and logical sequencing facilitated ease of interpretation, allowing viewers to extract insights without visual overload.
- Visual Hierarchy: The design approach was informed by theoretical frameworks in visual aesthetics. Binkley (1970) and Carroll (2001) assert that the effectiveness of visualization depends not only on data accuracy but also on formal elements—such as color, structure, and spatial arrangement—that shape user perception. Walton (1970) further emphasizes that perceived importance is influenced by the categorical context of information. Applying these insights, the dashboard layout prioritized critical data points while avoiding clutter, enhancing user engagement and comprehension.

Ethical Reflection:

- Transparency and Accuracy: Visual representations adhered strictly to the underlying data, avoiding misleading scales or omission of data points. Profit trends and regional disparities were communicated clearly and without distortion.
- Data Integrity: The visualizations maintained ethical standards by ensuring that no graphical manipulation or selective reporting compromised the credibility of the findings. Emphasis was placed on accurate representation to support sound business judgment.

3.2 Practice – Tool-Driven Redesign with Google Colab

Figure 8. Flowchart of Redesign Process in Colab

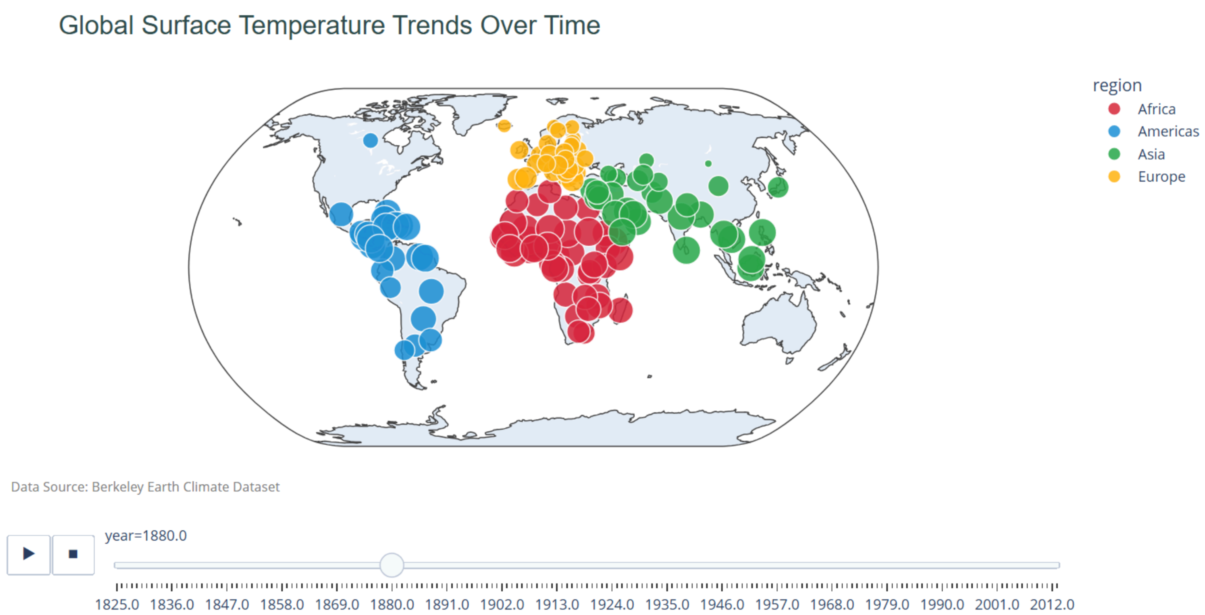
Source: Created by Canva

The final redesign project utilizes Google Colab to create an interactive map visualizing global temperature data. This map enables users to explore temporal trends in temperature changes through a time slider and to access additional details by hovering over specific countries. Continent and country labels, particularly for high-impact regions such as China, the United States, and India, were added to enhance geographic focus. Fonts and color schemes were customized to improve visual clarity and legibility. Temperature values are encoded using both an intuitive color gradient and marker size, facilitating a clearer understanding of global warming patterns.

The interactive capabilities of Plotly significantly shaped the visualization strategy, allowing for dynamic engagement with the dataset. The use of a "natural earth" projection ensured a geographically balanced representation of the world, while layout adjustments prioritized readability. Overall, this design approach enhances the accessibility, interpretability, and communicative power of complex climate data.

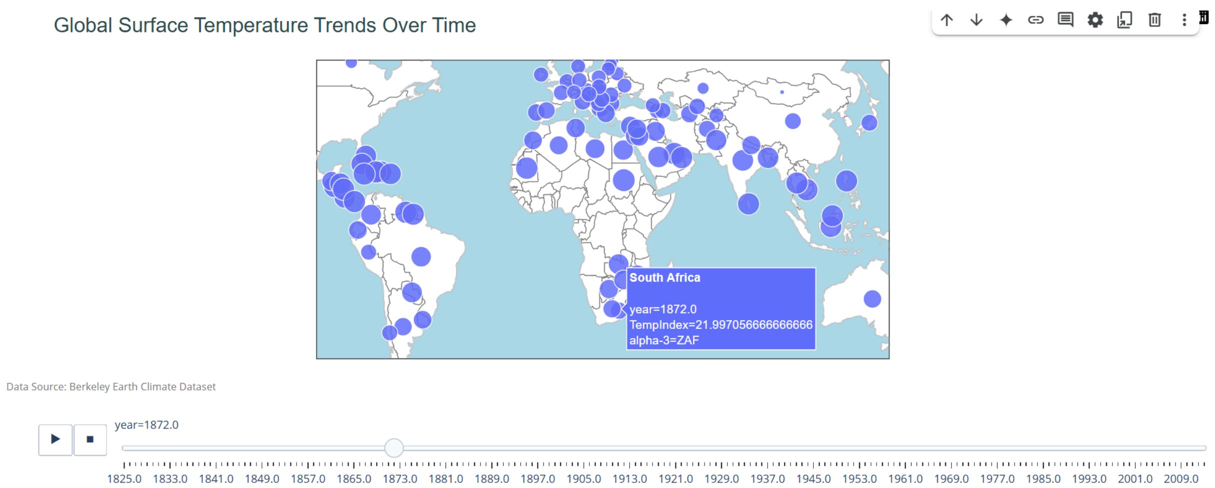
4. Practice – Final Redesign _Address Speaker's Feedback

Figure 9.1 Global Surface Temperature Trends Over Time_Original



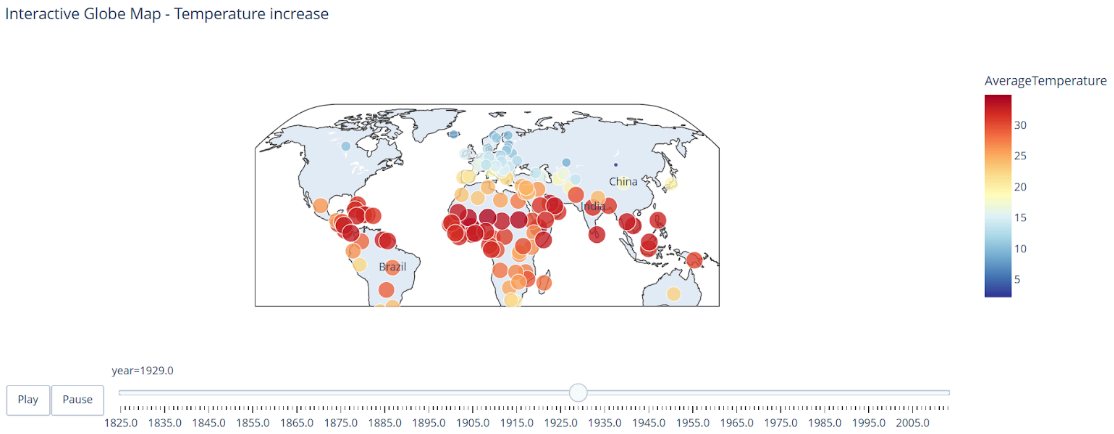
Source: Created by Colab

Figure 9.2 Global Surface Temperature Trends Over Time_Revised after reflection



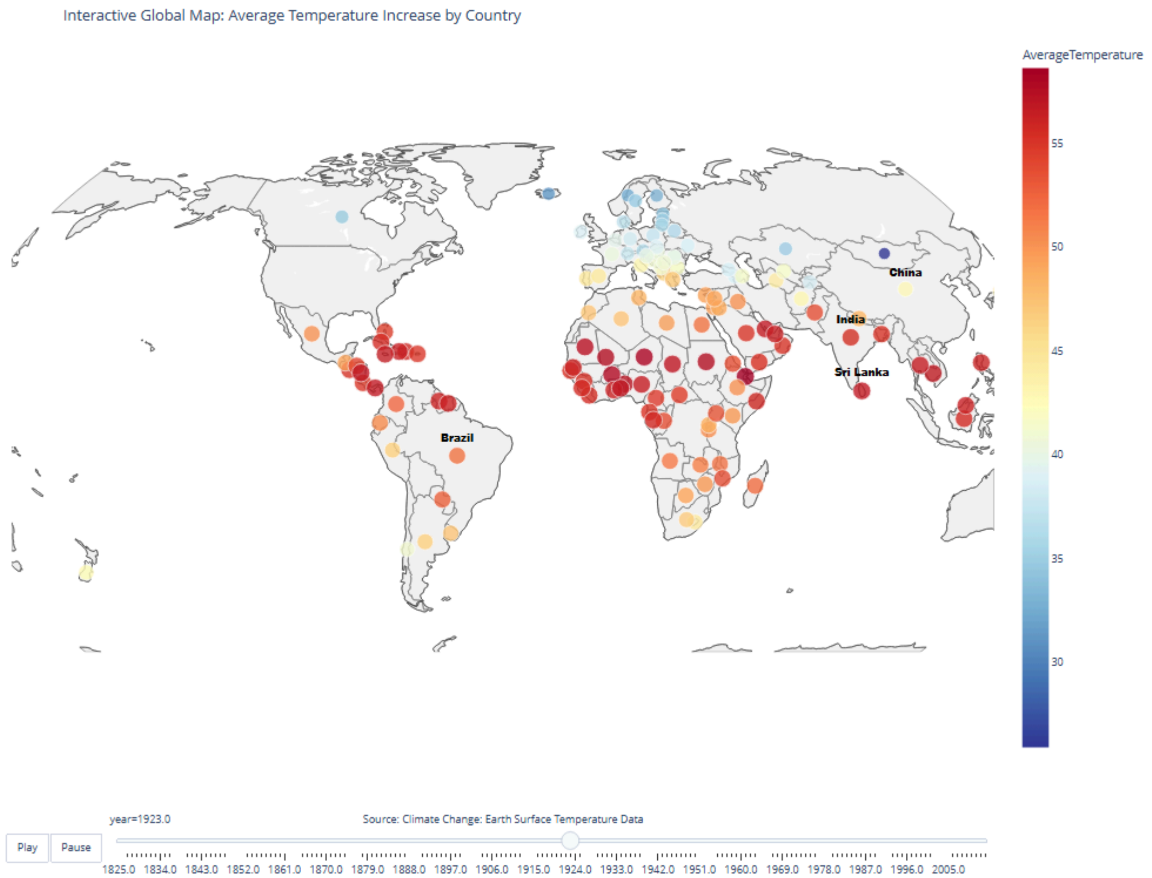
Source: Created by Colab

Figure 10.1 Interactive Global Map: Average Temperature Increase by Country_Original



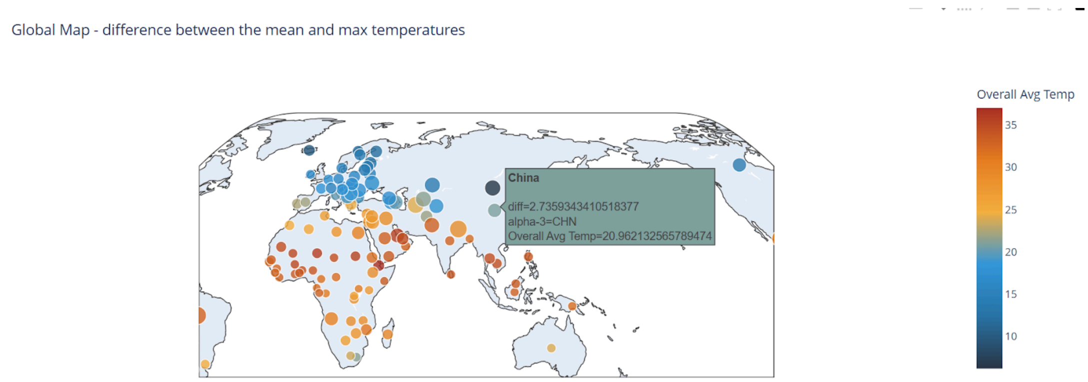
Source: Created by Colab

Figure 10.2 Interactive Global Map: Average Temperature Increase by Country_Revised after reflection



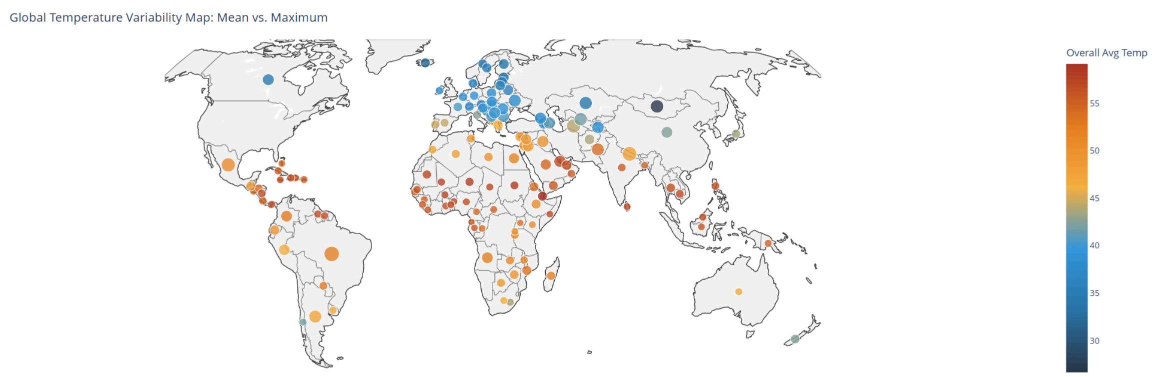
Source: Created by Colab

Figure 11.1 Global Temperature Variability Map: Mean vs. Maximum_Original



Source: Created by Colab

Figure 11.2 Global Temperature Variability Map: Mean vs. Maximum_Revised after reflection



Source: Created by Colab

In this redesign project, I utilized Plotly in Google Colab to create three interactive global temperature visualizations. My objective was to move beyond static visuals and leverage Python-based tools to enhance user engagement and analytical depth. Plotly was selected for its ability to generate highly interactive and dynamic visuals directly within a browser-based environment. Compared to traditional libraries like Matplotlib or Seaborn, Plotly provides built-in support for animations, tooltips, and geographic projections, making it especially suitable for time-series and spatial data. These capabilities significantly influenced my redesign strategy, allowing me to build intuitive visualizations that invite users to explore trends and patterns over time and across locations.

The new visuals improve upon the original in several key ways. First, there is enhanced temporal interactivity: users can now interact with a time slider, making it easier to track temperature changes over the years rather than viewing a static snapshot. Regarding geographic context, I added continent and key country labels (e.g., China, the United States, India, Brazil) with customized fonts to improve clarity and draw attention to regions with a significant global impact. Additionally, I refined the color encoding: temperature values are now mapped using a more intuitive color scale and marker size, making global warming trends more visually apparent. I also standardized the temperature values to avoid negative sizes, ensuring better visual balance. Finally, projection and layout enhancements were made by using the "natural earth" projection for a more realistic and balanced world view. The figure layout and annotations were also adjusted to improve readability and provide clearer data source traceability.

The visualization I created aligns with several important data principles, such as the FAIR Data Principles (Wilkinson et al., 2016), ensuring that the dataset is both accessible and reusable for users. By providing clear geographic labels and an interactive time slider, users can easily explore the data, promoting accessibility and user engagement. The visualization also respects the OECD Data Governance Framework (OECD, 2019) by ensuring data transparency and traceability through annotations that cite the data source. The dataset is made findable and interpretable, and the clear visual encoding ensures that users can interact with the data in a way that respects ethical and responsible data practices, as outlined in the DMBOK (DMBOK, 2017).

Data & Code Availability Statement

The dataset used for the global temperature visualizations is publicly available from [https://www.kaggle.com/datasets/berkeleyearth/climate-change-earth-surface-temperature-data] and was accessed in [April, 2025]. All code used to generate the visualizations is written in Colab and is available upon request or via the project's GitHub repository: [YifeiCathyYang/Redesign-Project: INFO 301 Information Visualization Redesign Project].

- **Dataset: Climate Change: Earth Surface Temperature Data**
  - https://www.kaggle.com/datasets/berkeleyearth/climate-change-earth-surface-temperature-data
- **GitHub Redesign repository**
  - YifeiCathyYang/Redesign-Project: INFO 301 Information Visualization Redesign Project

Statement of Contribution to Sustainable Development Goals (SDGs)

This project contributes to the following Sustainable Development Goals (SDGs).In alignment with Duke Kunshan University's vision for responsible innovation and global impact, the project also emphasizes sustainability, equity in information access, and environmental education.

SDG 3 Good Health and Well-Being: By critically redesigning climate and health-related visualizations, this project enhances public awareness of environmental risks (e.g., heat stress exposure) that directly impact human health. Improved data storytelling empowers stakeholders, communities, and policymakers to make better-informed health and safety decisions, ultimately supporting preventive measures against environmental health risks.

Figure 12. SDG 3 Good Health and Well-Being

Source: The United Nations

SDG 11 Sustainable Cities and Communities: The visualization improvements promote climate resilience by making complex environmental data more accessible and actionable. By helping communities understand geographical vulnerabilities and temperature changes over time, this work supports the development of sustainable, disaster-resilient cities and communities.

Figure 13. SDG 11 Sustainable Cities and Communities

Source: The United Nations

Future Research Direction on Digital Humanities (with Field Trip Integration)

The field trip to the Zhouzhuang Mystery of Life Museum inspired several future research directions at the intersection of digital humanities, biodiversity, and community-based learning. One promising direction is the application of 3D modeling technologies to enhance audience engagement and scientific understanding. For instance, exhibits such as the liver cell nucleus and alcoholic cirrhosis tissues could be reconstructed into interactive 3D models, allowing visitors to explore microscopic health impacts in a more vivid and tangible way (see Figure 14). This approach would not only bridge biological science with digital humanities but also promote deeper public awareness of health and environmental issues.

Figure 14. Nucleus of Liver Cell (TEM)

Source: The Zhouzhuang Mystery of Life Museum

In addition, the exhibit documenting the collaborative effort behind animal plastination highlighted the importance of teamwork and craftsmanship in scientific projects (see Figure 15). Future visualization projects could integrate behind-the-scenes storytelling, bringing to light the human dedication, ethical considerations, and scientific rigor that drive such endeavors. Interactive narratives that weave together scientific aesthetics and ethics could foster a stronger emotional connection between audiences and the scientific content they encounter.

Moreover, the museum experience underscored the potential for community-based biodiversity education. Future projects could involve local citizens in documenting regional biodiversity through digital platforms and visualizing these contributions via interactive tools such as augmented reality applications. This would empower communities to actively participate in environmental preservation and education initiatives. By integrating scientific storytelling, aesthetics, technology, and ethics, future digital humanities projects can create immersive, impactful experiences that inspire greater understanding and stewardship of both human health and the natural world.

Figure 15. Collaborative Effort Behind Animal Plastination

Source: The Zhouzhuang Mystery of Life Museum

Acknowledgement

This project has greatly benefited from the discussions at the *Digital Technology for Sustainability Symposium* at Duke Kunshan University on April 18. I am especially grateful to Professor Fan Liang for his valuable insights, which helped refine my work, and to the conference organizers, Professors Luyao Zhang, Fan Liang, and Charles Chang, for making the symposium possible.

I would also like to express my sincere gratitude to Professor Luyao Zhang for her insightful instruction throughout the INFOSCI 301 course—her guidance was instrumental in shaping the direction of this project. Special thanks to guest speakers Dongping Liu

(Amazon) and David Schaaf (Saarland University) for their thoughtful contributions, which significantly informed the redesign process. Finally, I extend my appreciation to my classmates in INFOSCI 301 for their constructive feedback and collaboration throughout the semester.

References

Amazon Web Services. "Leveraging the Cloud for Rapid Climate Risk Assessments." AWS Public Sector Blog, July 27, 2022. Leveraging the cloud for rapid climate risk assessments | AWS Public Sector Blog

Amazon Web Services, Inc. "Business Intelligence Tools - Amazon QuickSight - AWS." Accessed March 23, 2025. Amazon Web Services, Inc. "Business Intelligence Tools - Amazon QuickSight - AWS," n.d. https://aws.amazon.com/quicksight/.

Binkley, T., 1970, "Piece: Contra Aesthetics," *The Journal of Aesthetics and Art Criticism*, 35: 265–277.

Carroll, N., 2000, "Art and the Domain of the Aesthetic," *The British Journal of Aesthetics*, 40: 191–208.

Data Management Association International. (2017). *Data management body of* knowledge (DMBOK) (2nd ed.). Technics Publications. https://www.dama.org/cpages/home

Lee, Doris & Tang, Dixin & Agarwal, Kunal & Boonmark, Thyne & ChenCaitlyn & Kang, Jake & Mukhopadhyay, Ujjaini & Song, Jerry & Yong, Micah & Hearst, Marti & Parameswaran, Aditya. (2021). Lux: always on visualization recommendations for exploratory data framework flows.Proceedings of the VLDB Endowment.15.727-738.10.14778/3494124.3494151.

Map Library. (n.d.). *12 data visualization techniques that transform modern maps*. Map Library. Retrieved March 22, 2025, from https://www.maplibrary.org/1594/exploring-data-visualization-techniques-in-maps/

Munzner, T. (2014). Visualization Analysis and Design (1st ed.). A K Peters/CRC Press. https://doi.org/10.1201/b17511

Organisation for Economic Co-operation and Development (OECD). (2019). *OECD data governance framework*. OECD Publishing.

Walton, K.L., 1970, "Categories of Art," *The Philosophical Review*, 79 (3): 334–367.

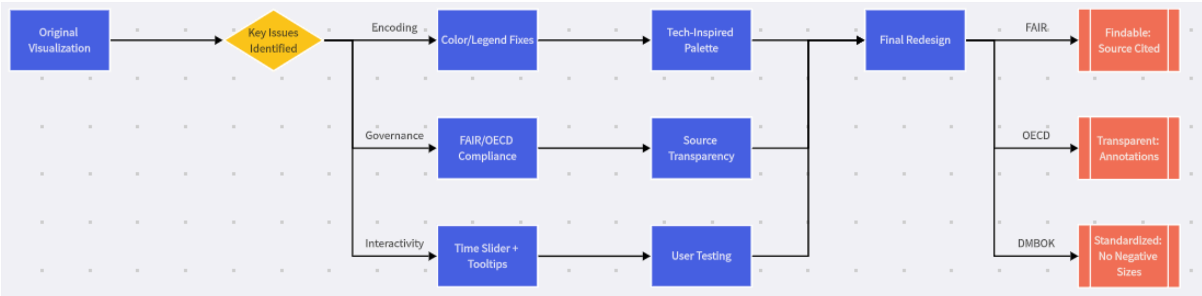
Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.1

Wongsuphasawat, K., Moritz, D., Anand, A., Mackinlay, J., Howe, B., & Heer, J. (2016). Voyager: Exploratory Analysis via Faceted Browsing of Visualization Recommendations. IEEE Transactions on Visualization and Computer Graphics, 22(1), 649-658.

Yu, Yue & Leixian, Shen & Long, Fei & Qu, Huamin & Chen, Hao. (2024). PyGWalker: On-the-fly Assistant for Exploratory Visual Data Analysis. 10.48550/arXiv.2406.11637.

5. Appendix: Additional Flowcharts

Flowchart 1: Redesign Workflow & Data Principles Check



Flowchart 2: Implementation Process in Colab