PictoLens: Gaze-Driven Interaction Technique for Layered Data Visualization Exploration

SARAH MAKAREM

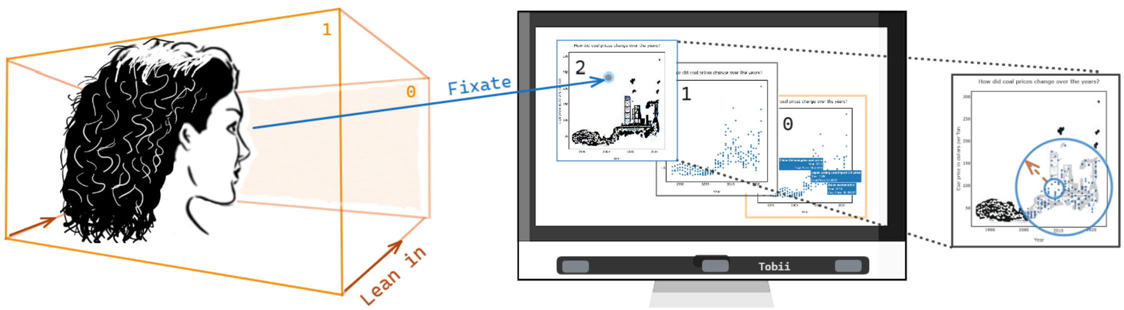


Fig. 1. PictoLens enables intuitive and dynamic exploration of layered data visualisations using gaze and head motion gestures. The figure shows how a user can fixate on a point on the pictogram layer (layer 2) to reveal a more detailed scatterplot visualisation (layer 1). A lean-in action further zooms in, revealing the most detailed layer (layer 0) with annotated data points. This system uses progressive disclosure to present and explore complex data in a natural way. The gaze serves as a pointer, and head movement controls the level of detail, enabling seamless and hands-free interaction.

PictoLens is a novel gaze-based interaction technique for exploring layered data visualizations through progressive disclosure. The system uses real-time gaze data to implement a point-and-click interaction model. Through intuitive gestures such as 'Gaze and Fixate' and 'Gaze and Lean In,' users can seamlessly interact with three representations of the data: an AI-generated pictograph, a scatter-plot visualization, and an annotated scatter-plot visualization. This hands-free and voice-free interaction technique addresses key challenges of traditional data exploration, such as long dwell times and the Midas Touch problem. PictoLens uses intuitive metaphors from everyday gestures: the gaze serves as a pointer, moving the visualization lens. Fixating the gaze at a point on the pictograph unlocks a finer data representation, while leaning forward reveals the most granular, detailed visualization layer with annotations. We present PictoLens' design and implementation to demonstrate its potential as an immersive analytics tool and interaction technique.

CCS Concepts: • **Human-centered computing → Human computer interaction (HCI)**; **Interaction techniques**; *Visualization systems and tools.*

Additional Key Words and Phrases: Gaze-Based Interaction, Human-Computer Interaction (HCI), Data Visualization, Interaction Techniques, Accessibility, Hands-Free Interaction, Progressive disclosure

ACM Reference Format:

Sarah Makarem. 2026. PictoLens: Gaze-Driven Interaction Technique for Layered Data Visualization Exploration. 1, 1 (January 2026), 7 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

Author's Contact Information: Sarah Makarem, sarah.makarem@kit.edu.

Manuscript submitted to ACM

1 Introduction and Related Work

Infographics are a widely used method for visually representing data and making complex information easy to understand [11]. They use visual elements to represent data in an appealing and memorable way. Studies have shown that effective visual embellishments and decorations can significantly improve the memorability and understanding of data visualizations [1–3, 5, 6, 14], especially in environments where the viewer's attention may be divided, such as meetings and lectures [4]. For these types of data representations, traditional methods for data exploration, such as mouse or touch interactions, often reach their limits when it comes to intuitive and hands-free exploration of complex datasets. Gaze-based interaction techniques offer a promising alternative.

We present PictoLens, a gaze-based interactive system that augments gaze-based gestures to interact with layered data visualizations for dynamic exploration. Gaze-based interaction often suffers from the so-called Midas Touch problem [17], where gaze fixation for a certain duration can unintentionally trigger an action. Relying on dwell times for visualization lenses slows the execution of commands, disrupting the smooth progression of the workflow [16]. PictoLens solves the Midas Touch problem by using intentional actions such as 'gaze and fixate' and 'gaze and lean in.' This enables natural, real-time transitions between three visualization layers: pictograph, bare data visualization, and annotated data visualization. PictoLens uses intuitive metaphors from everyday life to design the interaction technique. As a special case of the point-and-click interaction model, gaze serves as a pointer that moves the visualization lens. For explicit interaction, a second modality is used. Fixating the gaze on one point allows insight into an underlying layer of data visualization, while leaning forward enlarges the lens to reveal increasingly detailed data. This intuitive gesture mimics our natural behavior of approaching objects to see finer details. The interplay of pictographs as visual metaphors and natural eye movement enables a focused and detailed exploration of the data. Annotations are displayed only for the data point in focus, similar to foveal vision, where the highest resolution is achieved in the center of the foveal region. Technically, PictoLens is implemented using a Tobii Eye Tracker and a Flask-based server architecture to process real-time gaze and head tracking data. The system recognizes gestures and dynamically updates the visualization in a web-based interface. The aim is to provide low-latency interaction, allowing users to seamlessly navigate through the data visualization layers and dynamically unveil data insights.

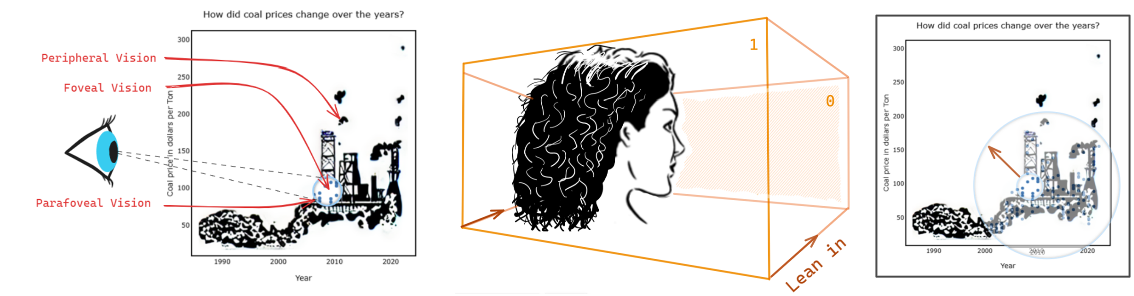
2 Methodology

2.1 Pictograph Generation

Three data representations are generated from the same dataset:

(1) **Layer 2: Pictograph**: An abstract, high-level representation of the data to spark user interest.
(2) **Layer 1: Bare Data Visualization**: A detailed scatter plot based on the same data as visualized in layer 2.
(3) **Layer 0: Annotated Data Visualization**: A detailed scatter plot enriched with labels based on the same data.

The generation workflow starts with creating a standard scatter plot from the raw data. The pictograph is then automatically generated using the generation pipeline from PictographAI [10]. The necessary context information is collected from the news article and fed into a Langchain LLM agent [9], which extracts a suitable prompt of one or two words. In the following step, a text-guided image-to-image in-painting Stable Diffusion model [8] uses the text input from the LLM agent to generate a context-relevant pictographic representation of the scatter plot. Afterwards, post-processing operations are performed to overlay the pictograph on the original chart axes. This step ensures that the generated pictograph maintains the accuracy of the standard data visualization.



(a) PictoLens metaphor of the human visual field.

(b) Visualization lens dynamically grows as the user leans in, revealing more granular data representation

Fig. 2. PictoLens Interaction Metaphors.

2.2 Interaction Design

Just as pictographs serve as visual metaphors for data visualizations, our approach draws inspiration from intuitive interaction metaphors derived from natural human behaviors to shape this interaction technique.

2.2.1 The Human Visual Field. The human visual field spans approximately 135 x 220 degrees and can be divided into three main regions: foveal, parafoveal, and peripheral [7]. Although the foveal region accounts for less than 1% of our visual field, it plays a crucial role in processing detailed visual information. In PictoLens, the foveal region serves as a metaphor for the central part of the lens, where users can examine sharper and more detailed data representations.

The parafoveal region, with slightly lower visual sharpness, aids in perceiving broader details and providing context. In PictoLens, this corresponds to the lens border and inner shadow, enabling users to perceive depth and contextual relationships. The peripheral region, which extends to the outer edges of the visual field, has the poorest sharpness but provides essential context about the surroundings. In PictoLens, this region metaphorically represents the pictographic visualization surrounding the primary point of interest, as shown in Figure 2a.

Eye movements play a critical role in scanning the environment, directing the foveal region toward areas of interest. Similarly, PictoLens allows users to explore focused and granular representations of data visualizations simply by moving their gaze.

This concept is further reflected in Layer 2 of PictoLens, where annotations appear only for data points users fixate on, corresponding to the foveal region. By optimizing visual processing for the area where the gaze is focused, PictoLens captures the highest-resolution data for that point, mimicking the brain's ability to process detailed visual information selectively. This mechanism is replicated through dynamic annotations, enabling users to unlock the most granular level of data representation: individual data points.

2.2.2 Layered Interaction. Gaze and Fixate, along with Gaze and Lean In, are intuitive gestures modeled after the point-and-click interaction model. In PictoLens, gaze acts as a pointer, moving the visualization lens. To trigger an intentional action, a secondary modality is required. In the pictograph layer, the user must fixate on a dot for a minimum dwell duration of 600 ms, which has been found to be the most efficient for eye-computer interaction interface design. This duration exceeds the typical fixation range for learning in the context of information visualization (300–500 ms) [12, 13]. Upon fixation, the dot expands into a circular lens, offering a glimpse into Layer 1: the scatter plot. Once the

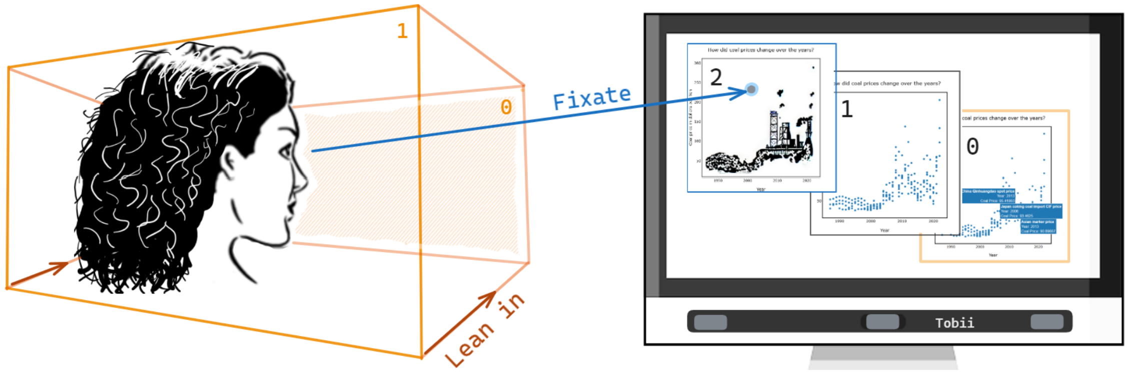


Fig. 3. Interaction model with layered visualizations (Layers 0, 1, 2) corresponding to user head movement (planes 1, 0) to transition across visualization layers.

lens appears, the user can freely gaze around to explore the visualization on Layer 1. The fixation radius is calculated based on a fixation threshold of 2° of the visual angle and a visualization size of 512x512 pixels, resulting in a radius of 66 pixels. The term Gaze and Fixate here is shortened from Gaze, Fixate and Dwell.

When the user becomes intrigued and leans closer to the screen, PictoLens mirrors this behavior. As the user moves from Plane 1 to Plane 0, PictoLens expands the visualization lens and transitions from Layer 1 to Layer 0, unveiling the innermost layer of the data visualization. On Layer 0, the user's gaze triggers automatic annotations for the point of fixation and the two closest data points, enabling more detailed exploration and interaction with the data.

2.2.3 Zooming Interaction. Lens growth is synchronized with user movement as they lean their head forward, as shown in Figure 2b. The metaphor for leaning in is akin to zooming in. As users move closer to collect more information, the lens expands to reveal increasingly granular data from the visualization layer beneath. This intuitive gesture mimics the natural behavior of physically approaching objects to see finer details.

2.2.4 Progressive Disclosure. PictoLens uses a layered approach to visualizations, progressively revealing more detailed information as the user interacts with the system. This approach helps manage the complexity of the data, supports an exploratory workflow, and minimizes cognitive load [15].

2.2.5 Dwell Time. PictoLens utilizes dwell interactions when the user fixates on a dot for a minimum dwell time of 600 ms on Layer 2. The same minimum dwell time is applied on Layer 0 to trigger chart annotations. To avoid the Midas Touch problem on Layer 0, the gaze data stream is paused after each annotation appears, allowing the user sufficient time to read the annotation.

3 Technical Realization

PictoLens utilizes a **Tobii Pro Spark Eye Tracker** for real-time gaze tracking and head movement detection, along with a display setup for rendering multi-layered visualizations in real time. On the software side, PictoLens **Real-Time Data Processing Pipeline** integrates several components to enable gaze-driven exploration of layered data visualizations. The key components of the pipeline are described below:

Tobii Pro Eye Tracking Manager: Used for calibrating the eye tracker to the user's eyes. Real-time data communication is handled by the **Tobii Software Development Kit**, which collects: **Gaze Position,** indicating the real-time

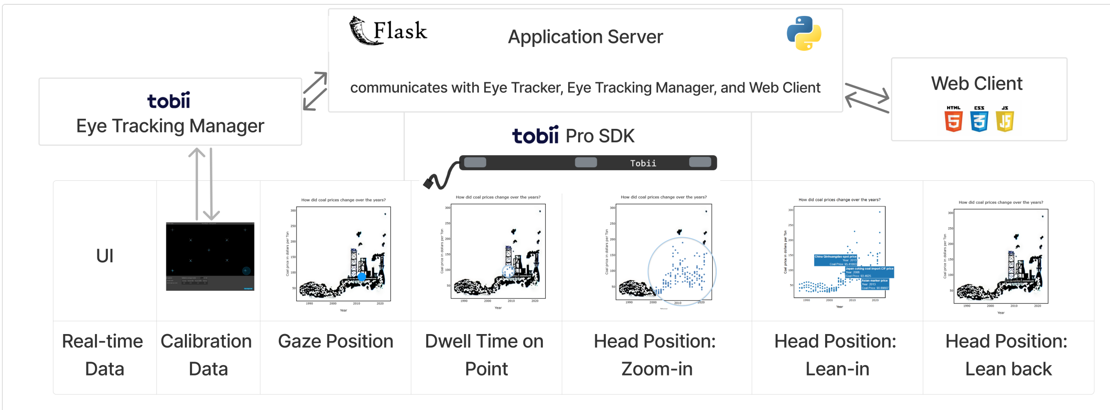


Fig. 4. Interaction flow mapping real-time data to user interface

location on the screen reflected in the dot, lens, or annotations; **Dwell Time,** measuring how long the user's gaze remains fixed on a specific location, expanding the lens after 600 ms within the same 66 Pixels; and **Head Position,** tracking movements such as zooming in, leaning in, or leaning back.

Server: A Flask-based server facilitates communication between the eye tracker, the SDK, and the web client. It identifies interaction gestures such as: *Gaze and Fixate,* expanding a dot into a lens to explore the next layer; *Gaze and Zoom In,* expanding the lens to reveal more detailed layers; *Gaze and Lean In,* accessing the lowest layer with annotations; and *Gaze and Lean Back,* returning to higher-level layers.

Web Client: The processed data updates a web-based interface built with HTML, CSS, and JavaScript, enabling real-time interaction. The client handles: *Layer Transitions,* switching between the pictograph (Layer 2), bare visualization (Layer 1), and annotated visualization (Layer 0); and *Annotations,* dynamically displaying contextual labels based on gaze position by programmatically triggering a hover action.

4 Demonstration

During the demonstration of the PictoLens system, users first need to be briefly calibrated to the eye tracker. They can then freely interact with the various data and explore different forms of presentation. PictoLens enables intuitive and dynamic exploration of data visualisations using gaze and head movement gestures. The system is designed to allow users to switch between the various levels seamlessly and intuitively by using natural gestures. The progressive disclosure of data makes it possible to explore complex data sets in a clear and controlled manner. PictoLens enables hands-free interaction with various data sets and their representations.

5 Discussion

During the design phase, many requirements and usability considerations were carefully evaluated and incorporated into the development process. However, there are inherent limitations as well as opportunities for future improvements, which we discuss below.

5.1 Limitations

Range of head motion. PictoLens relies on estimating gaze position as the user performs head movement gestures, but the calibration of the eye tracker is conducted when the head is in a fixed position. This can result in some inaccuracy when the user is leaned in or leaned back, depending on the head position used during calibration.

Dependency on accurate gaze-tracking hardware. PictoLens currently depends on specialized hardware, such as the Tobii Eye Tracker, for real-time gaze tracking and head movement detection. The accuracy and performance of the system are directly tied to the quality of the hardware used.

Learning curve for first-time users. While PictoLens is designed to be intuitive, first-time users may still need to familiarize themselves with the system's interaction paradigms, such as using gaze and head movements to navigate through visualization layers. Although the system offers a quick setup and calibration process, we have not yet assessed the learning curve for the target user groups. This could necessitate training or guidance for new users, particularly in environments where rapid adoption is critical.

5.2 Advantages and Opportunities

Intuitive, hands-free interaction. PictoLens leverages gaze-tracking technology to enable users to interact with visualizations without the need for traditional input devices such as a mouse or keyboard. This allows for a more natural and intuitive interaction, mirroring how we explore our physical environment with our eyes. This hands-free approach is particularly beneficial in scenarios where manual operation is impractical or impossible, such as in medical settings, industrial environments, or for users with physical disabilities.

Progressive disclosure through layered visualizations. PictoLens employs a novel layered approach to visualizations, progressively revealing more detailed information as the user interacts with the system. This design helps manage the complexity of data by presenting it incrementally, reducing cognitive load and enhancing user engagement. The structured and digestible presentation of information makes this approach ideal for many complex, data-driven applications.

Broad applications. The hands-free and intuitive nature of PictoLens makes it highly versatile across various domains. In accessibility, it can support users with motor impairments by providing an alternative interaction method. In education, it offers an engaging way to explore complex datasets, fostering deeper understanding. In AR/VR, it enhances immersive experiences by enabling natural, gaze-driven interaction with virtual environments. This adaptability positions PictoLens as a valuable tool for a wide range of use cases.

References

[1] Tiffany Andry, Christophe Hurter, François Lambotte, Pierre Fastrez, and Alexandru Telea. 2021. Interpreting the effect of embellishment on chart visualizations. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.

[2] Scott Bateman, Regan L Mandryk, Carl Gutwin, Aaron Genest, David McDine, and Christopher Brooks. 2010. Useful junk? The effects of visual embellishment on comprehension and memorability of charts. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2573–2582.

[3] Scott Bateman, Regan L. Mandryk, Carl Gutwin, Aaron Genest, David McDine, and Christopher Brooks. 2010. Useful junk? the effects of visual embellishment on comprehension and memorability of charts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 2573–2582. doi:10.1145/1753326.1753716

[4] Rita Borgo, Alfie Abdul-Rahman, Farhan Mohamed, Philip W. Grant, Irene Reppa, Luciano Floridi, and Min Chen. 2012. An Empirical Study on Using Visual Embellishments in Visualization. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2759–2768. doi:10.1109/TVCG.2012.197

[5] Michelle A. Borkin, Azalea A. Vo, Zoya Bylinskii, Phillip Isola, Shashank Sunkavalli, Aude Oliva, and Hanspeter Pfister. 2013. What Makes a Visualization Memorable? *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2306–2315. doi:10.1109/TVCG.2013.234

[6] William S. Cleveland and Robert McGill. 1984. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *J. Amer. Statist. Assoc.* 79, 387 (1984), 531–554. http://www.jstor.org/stable/2288400

[7] David C. Van Essen and Charles H. Anderson. 1995. Information Processing Strategies and Pathways in the Primate Visual System . https://api.semanticscholar.org/CorpusID:18001786

[8] Hugging Face. 2024. Inpainting. https://huggingface.co/docs/diffusers/v0.29.2/en/using-diffusers/inpaint. Accessed: 2024-07-05.

[9] LangChain. 2024. Using Language Models. https://python.langchain.com/v0.2/docs/tutorials/agents/#using-language-models. Accessed: 2024-07-05.

[10] Sarah Makarem, Tobias Röddiger, Till Riedel, and Michael Beigl. 2024. PictographAI: Interactive Generation of Stylized Pictographs for Presentations. In *Mensch und Computer 2024 - Workshopband, Karlsruhe, Germany, September 1-4, 2024*. Gesellschaft für Informatik e.V. doi:10.18420/MUC2024-MCI-DEMO-305

[11] Thomas Moritz and Thomas Moritz. 2019. Prinzipien Visueller Wahrnehmung als Voraussetzung Visueller Kommunikation. *Screenografie kompakt: Der immersive Bildraum grafischer Benutzeroberflächen* (2019), 37–72.

[12] Shivsevak Negi and Ritayan Mitra. 2020. Fixation duration and the learning process: an eye tracking study with subtitled videos. *Journal of Eye Movement Research* 13 (08 2020). doi:10.16910/jemr.13.6.1

[13] Ya-feng Niu, Yue Gao, Cheng-qi Xue, Ya-ting Zhang, and Li-xin Yang. 2020. Improving Eye-Computer Interaction Interface Design: Ergonomic Investigations of the Optimum Target Size and Gaze-triggering Dwell Time. *Journal of Eye Movement Research* 12 (09 2020). doi:10.16910/jemr.12.3.8

[14] Drew Skau, Lane Harrison, and Robert Kosara. 2015. An evaluation of the impact of visual embellishments in bar charts. In *Computer Graphics Forum*, Vol. 34. Wiley Online Library, 221–230.

[15] Aaron Springer and Steve Whittaker. 2019. Progressive disclosure: empirically motivated approaches to designing effective transparency. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (Marina del Ray, California) *(IUI '19)*. Association for Computing Machinery, New York, NY, USA, 107–120. doi:10.1145/3301275.3302322

[16] Christian Tominski, S. Gladisch, Ulrike Kister, Raimund Dachselt, and H. Schumann. 2016. Interactive Lenses for Visualization: An Extended Survey. *Computer Graphics Forum* 36 (05 2016). doi:10.1111/cgf.12871

[17] Hana Vrzakova and Roman Bednarik. 2013. That's not norma (n/l) a detailed analysis of midas touch in gaze-based problem-solving. In *CHI'13 Extended Abstracts on human Factors in Computing Systems*. 85–90.