

# Lumos: Increasing Awareness of Analytic Behavior during Visual Data Analysis

Arpit Narechania, Adam Coscia, Emily Wall, Alex Endert

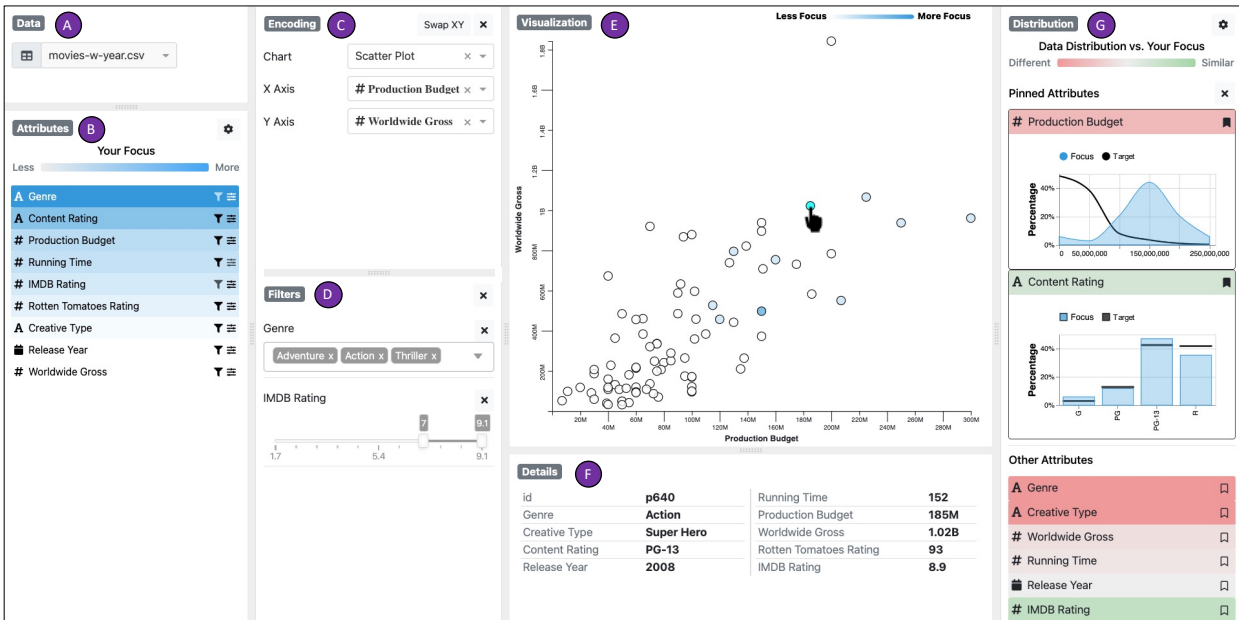


Fig. 1: The Lumos user interface includes traditional visual data analysis functions ((A) Data Panel, (B) Attributes Panel, (C) Encoding Panel, (D) Filter Panel), and shows analytic behavior as in-situ and ex-situ interaction traces in the (B) Attributes Panel, (E) Visualization Canvas, (F) Details View, and (G) Distribution Panel as the relation between a target distribution (e.g., the underlying data) and the user’s analytic behavior.

**Abstract**—Visual data analysis tools provide people with the agency and flexibility to explore data using a variety of interactive functionalities. However, this flexibility may introduce potential consequences in situations where users unknowingly overemphasize or underemphasize specific subsets of the data or attribute space they are analyzing. For example, users may overemphasize specific attributes and/or their values (e.g., Gender is always encoded on the X axis), underemphasize others (e.g., Religion is never encoded), ignore a subset of the data (e.g., older people are filtered out), etc. In response, we present Lumos, a visual data analysis tool that captures and shows the interaction history with data to increase awareness of such analytic behaviors. Using in-situ (at the place of interaction) and ex-situ (in an external view) visualization techniques, Lumos provides real-time feedback to users for them to reflect on their activities. For example, Lumos highlights datapoints that have been previously examined in the same visualization (in-situ) and also overlays them on the underlying data distribution (i.e., baseline distribution) in a separate visualization (ex-situ). Through a user study with 24 participants, we investigate how Lumos helps users’ data exploration and decision-making processes. We found that Lumos increases users’ awareness of visual data analysis practices in real-time, promoting reflection upon and acknowledgement of their intentions and potentially influencing subsequent interactions.

**Index Terms**—visual data analysis, interaction traces, analytic provenance, awareness, human bias

## 1 INTRODUCTION

Visualizations take advantage of people’s perception to facilitate intuitively understanding data. Interactive features of visualizations become critical when considering complex data, allowing people to progres-

sively refine visual representations of data. While it can aid in comprehension of large and complex data, certain patterns of interactivity can signal insular data analysis practices. Users may be unknowingly stuck inside an “echo chamber”, where their own unconscious biases may lead them to attend to certain parts of the data while neglecting others.

- Arpit Narechania, Adam Coscia, Alex Endert are with Georgia Institute of Technology. Emails: {arpitnarechania, acoscia6, endert}@gatech.edu. Emily Wall is with Emory University. Email: emily.wall@emory.edu \*work performed while at Northwestern University.

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxx/TVCG.201x.xxxxxxx

Unconscious biases can take many forms, some of which are relatively innocuous (i.e., personal preference for a particular genre of movie) while others can lead to costly incorrect decisions or engender harmful societal stereotypes. For instance, dark-skinned people are denied parole (racial bias) [3], women experience a variety of barriers to equity in C-suite promotions (gender bias) [20], ailing but younger people are denied optimal treatment (age bias) [29], etc. Even when these factors are not actively considered, they can implicitly affect the way people examine and process new information [16], and hence their

analytic behaviors.

Apart from implicit biases and stereotypes, there are other cognitive and perceptual biases that also influence people’s analytic behaviors. Cognitive biases, in particular, describe systematic errors that can result from the use of “fast and frugal” heuristics [14] to make decisions. Several biases have been shown to extend to tasks involving visualizations for decision making (e.g., [6, 8, 32, 35]). Yet, common visual data analysis tools such as Tableau and Microsoft Excel that help users see and understand their data do not report analytic behaviors that may correspond to such biases. Therefore we ask: how much can understanding data analysis and decision-making behaviors reduce the potentially negative influences of potential cognitive, perceptual, or societal biases, if users were simply more aware of these often unconscious factors?

We present Lumos, an analysis tool that visualizes interaction history with data (i.e., **interaction traces** [36]) to increase **awareness** of potential interaction biases that influence data analysis and decision making processes. Using *in-situ* and *ex-situ* visualization techniques, Lumos provides real-time feedback about a user’s **analytic behavior** for self-awareness and self-reflection to potentially change future course. For example, Lumos remembers and highlights datapoints that have been previously examined in the same visualization (in-situ) and overlays the interacted datapoints on the underlying data distribution in a separate visualization (ex-situ) for comparison. Furthermore, Lumos allows users to configure a custom target distribution to reflect decision making goals (e.g., a university admissions committee in a computer science department may define an analysis target of 60% female to promote increased gender diversity in the department, even if only 40% of applicants are female). We posit that Lumos can improve behaviors exhibited during data exploration and decision-making to help mitigate the dangers of human interaction biases affecting judgments and foster more transparent analysis processes.

The primary contributions of this work include (1) a technique to model users’ **analytic behavior** from interactions with the data, (2) a visual analysis tool, Lumos (<https://lumos-vis.github.io/>), that implements visualizing analytic behavior using interaction traces, (3) a series of scenarios that describe potential usage of Lumos, and (4) results from a user study to understand how Lumos helps users be more aware of their analytic behavior during visual data analysis.

## 2 RELATED WORK

### 2.1 Graphical History

We are limited by our memory’s capacity to remember and track our prior interactions with data in both amount and decay [21, 22] which creates a barrier to data exploration. Analyzing prior interactions with data in a visualization is one form of analytic provenance [24, 26] often used to infer about one’s analysis process (beyond analytic outcomes). Visualizing one’s prior interactions can take many forms and lead to shifts in a user’s analysis behavior. For instance, when users’ prior interactions with charts or data points are encoded (e.g., analogous to coloring previously visited hyperlinks on a webpage purple), people tend to interact with more of the data [12]. Similarly, when exploration history is shown in interactive network visualizations, users report inspiration for conducting further analysis and greater recall of their prior explorations [9]. Cleverly designed histories can also help users maintain contextual awareness of previously visited data when distortions are applied that would otherwise make contextual awareness a challenge (e.g., fisheye lens) [31]. More generally, Heer et al. summarize a design space of mechanisms for displaying interaction histories [18]. Given ongoing recognition within the visualization community that interaction histories can serve a wide range of purposes, recent tooling has been developed to support provenance tracking [7].

Graphical traces of user interactions have also been utilized in collaborative visualization settings, e.g., to facilitate coordination of multiple users by showing current selections and interactions as “coverage” of the data [5, 27] and in personalized integrated development environments (IDEs), e.g., Footsteps for VSCode<sup>1</sup> highlights the lines of code as they are edited by the user. Similarly, showing social information

“scents” on data visualization widgets (e.g., representing others’ interactions with radio buttons, sliders, etc.) leads users to make substantially more unique discoveries in the data [37].

Outside of the visualization domain, traces of prior interactions have long been applied in HCI contexts, including revisiting common regions of a page using scrollbar history [1], tracking interactions with documents by authors and readers [19], facilitating groupware coordination [17], tracking user focus while browsing a webpage using eye-tracking [23] and mouse-tracking [4] gear, etc. We take inspiration from this lineage of graphical representations of interaction history in our technique, **interaction traces**, expanded from [36] and described further in the next section.

### 2.2 Modeling User Behavior

Several metrics have been proposed in data visualization to characterize behaviors during data analysis with visualizations. For instance, Feng, Peck, and Harrison propose metrics to quantify *exploration uniqueness* and *exploration pacing* as users interact with points in a scatterplot [13]. Ottley, Garnett, and Wan use a hidden Markov model to capture user attention to predict clicks in a visualization [25]. Perhaps most relevant to the present work, Gotz, Sun, and Cao model and visualize the provenance of how a user’s subset selections of the data differ from the dataset as a whole [15]. We similarly model and visualize deviations of user behavior against a baseline; however, rather than focusing on explicit subset selections, we utilize a breadth of user interactions (including clicks, hovers, visualization configurations, filter configurations, etc.) as signals in our analytic behavior model.

Zhou et al. introduce a formal model of *focus* based on user interactions defined by (1) type of action and (2) focus of the action in the form of an additive model [38]. Wall et al. define metrics for quantifying *bias*, similarly based on (1) type of interaction and (2) object of interaction; however, these metrics differ in that they compare the observed focus to a baseline of “unbiased” behavior [34]. While these metrics have been used to quantify bias (e.g., anchoring), they are also more generally used to capture deviations from uniform behavior [33, 35]. Because it compares user behavior to a flexibly defined baseline, in this work we utilize the metrics by Wall et al. to model user behavior [34].

## 3 LUMOS

In this section we first define the terminologies used in the paper, enumerate our design goals, and describe the Lumos user interface.

**Interaction Logs.** Telemetry data of a user’s interactions (e.g., hovers) with the user interface elements (e.g., datapoints on a scatterplot).

**Analytic Behavior Model.** A model that quantifies user’s behavior (set of actions) from interaction logs (e.g., computing the distribution of interacted datapoints and attributes).

**Interaction Traces.** Visual feedback of the user’s analytic behavior in the user interface; this is shown by adding visual scents in two ways: *in-situ* (at the place of interaction) and *ex-situ* (in an external view).

**Awareness.** Knowing what is going on [11] or what has been done and found during the exploration process to perform effective reasoning [30]; in this context, knowledge gained from inspecting one’s analytic behavior via interaction traces.

### 3.1 Design Goals

Our development of Lumos was driven by **five** key design goals. We compiled these goals based on a combination of similar prior visual analysis tools [12, 15, 28], formative feedback from pilot studies, and our own hypotheses with respect to usability.

**DG1. Capture and present analytic behavior with attributes.** Overemphasis (or underemphasis) on specific attributes during data exploration may lead to unconscious biases (e.g., not interacting with a Gender attribute may practically result in a bias towards men if the dataset has more men than women). This goal translates to capturing user interactions with *attributes*, modeling analytic behavior, and showing interaction traces to increase awareness to influence changes in subsequent interactions.

**DG2. Capture and present analytic behavior with datapoints.** Overemphasis (or underemphasis) can also occur on specific values of

<sup>1</sup><https://marketplace.visualstudio.com/items?itemName=Wattenberger footsteps>

data (e.g., interacting mostly with a few top candidates for university admissions may come at the expense of neglecting other candidates). This translates to the same goal as **DG1** but at the datapoint-level.

**DG3. Facilitate configuring different target distributions.** Determining overemphasis (or underemphasis) on specific attributes or data requires comparing a user’s analytic behavior with a known target distribution (e.g., the underlying data) as a baseline. However, different domains, tasks, attributes, or social norms may call for different target distributions. This goal translates to allowing users to configure different target distributions to suit their requirements.

**DG4. Facilitate comparison between analytic behavior and a baseline distribution.** This goal translates to visualizing the user’s analytic behavior and the configured target distribution and quantifying the difference (or similarity) between the two distributions.

**DG5. Facilitate visual data exploration while showing awareness.** This goal ensures that visual data exploration usability is not sacrificed by the added awareness visualization techniques.

### 3.2 Quantifying Analytic Behavior

We quantify *analytic behavior* using (1) the attribute distribution (AD) metric [34] and (2) relative frequency of interactions with data and attributes. The AD metric, along a scale from 0 (no bias) to 1 (high bias), characterizes how a user’s interactive behavior deviates from expected behavior. By default, the system chooses a **proportional** baseline of expected behavior, wherein interactions with any given data point are equally likely, reflecting the true underlying distributions of attributes in the dataset. For example, if a user interacts primarily with PG-13 movies among a dataset of movies that contains predominantly G-rated movies, the AD metric for the Content Rating attribute will be high (more emphasis). If the user instead spent more time interacting with G-rated movies, proportional to the distributions in the dataset, the AD metric value for Content Rating would be low (less emphasis). Alternative baselines can be set using Lumos (Section 3.3.3).

### 3.3 User Interface

The Lumos user interface consists of the following views:

- (A) **Data Panel** shows the currently loaded dataset.
- (B) **Attribute Panel** shows a list of attributes in the dataset along with their datatypes: {Nominal (N; **A**), Quantitative (Q; **#**), Temporal (T; **📅**)} and buttons to apply a filter (**🔍**).
- (C) **Encoding Panel** shows UI controls (dropdowns) to create visualizations by specifying different encodings: {Chart Type, X Axis, Y Axis, Aggregation}. Lumos currently supports *four* visualization types: {*scatter plot*, *strip plot*, *bar chart*, *line chart*} and *five* aggregation types: {*count*, *sum*, *minimum*, *maximum*, *average*} depending on the attribute data type combinations.
- (D) **Filter Panel** shows UI controls (range sliders for {Q, T} and multiselect-dropdowns for {N} attributes) to filter data. Filters can be added by clicking on **🔍** in the Attribute Panel (B) (**DG5**).
- (E) **Visualization Canvas** renders the visualization based on the Encoding (C) and Filter (D) Panel specifications.
- (F) **Details View** shows additional information when the visualization elements (e.g., point, bar, strip) in (E) are interacted with. Hovering on a single datapoint (e.g., a strip in a strip plot) shows a list of all attribute values for the given datapoint (Figure 1F). Hovering on an aggregation of datapoints (e.g., a bar of a bar chart) shows a table of all datapoints that belong to that aggregation (the bar) with attributes as columns and values as rows (Figure 2).
- (G) **Distribution Panel** shows a list of attribute cards similar to the Attribute Panel where clicking on a card toggles open/close a visualization that overlays user’s interaction traces on datapoints (blue area) on the target distribution (black curve) (**DG3**).

In addition to these features, Lumos supports additional customization options (**⚙️**): (1) *Sort By* to sort attributes in the Attribute panel by their {Order-in-Dataset (default), Name, Datatype, Focus} (2) *Color Mode* to determine the normalization strategy to compute the interaction frequency: *Relative* (default) divides each value by the maximum of all values; *Absolute* divides each value by the sum of all values; *Binary* treats no interactions=0 and at least one interaction=1, (3) *Focus*

Details Release Year: 2008						
id	Genre	Creative Type	Content Rating	Release Year	Running Time	Prodi
p635	Action	Super Hero	PG-13	2008	126	
p637	Adventure	Super Hero	PG-13	2008	112	
p640	Action	Super Hero	PG-13	2008	152	
p659	Thriller	Contemporary Fiction	R	2008	129	
p670	Thriller	Dramatization	PG-13	2008	120	

Fig. 2: **Details View (Aggregate)**. Hovering on an aggregate visual element in the *Visualization Canvas* (e.g., a bar in a bar chart) updates the Details View with a table of movies with *Release Year=2008*; darker blue=greater emphasis toward that datapoint.

*Mode* that determines the Y-encoding in the Distribution Panel visualizations: *Percentage* (default) shows the percentage of Focus and Target distributions whereas *Raw* shows the absolute counts; and (4) *Color Scale* to choose different color scales (e.g., Sequential, Divergent).

Based on user interaction logs with attributes and datapoints, Lumos computes analytic behavior and presents real-time interaction traces back to the user. Analytic behavior and interaction traces are observed in multiple ways in the interface.

#### 3.3.1 In-situ Interaction Traces

**Visualization Canvas.** Lumos tracks user interactions with visual representations of datapoints (e.g., bars, lines, points, strips) and colors them on a white→blue scale based on the relative frequency of interactions, e.g., dark blue color represents more interactions (Figure 3) and white represents no interactions (**DG 2**). Lumos captures mouseover interactions as a proxy for modeling analytic behavior from interactions with datapoints. Lumos employs a heuristic to ignore mouseovers that are active less than a 350 milliseconds threshold, regarded as random, accidental, or unintentional.

An interaction with a unit visualization (e.g., hovering on a point in a scatter plot of *Running Time* and *Worldwide Gross*) is handled differently than an interaction with an aggregate visualization (e.g., hovering on a bar showing average *Running Time* of *Action* movies). In the former scenario, Lumos treats it as one *complete* interaction with the corresponding datapoint incrementing its interaction counter by 1. In the latter scenario, Lumos treats the interaction as a set of *partial* interactions with all constituent datapoints (e.g., all *Action* movies), incrementing their corresponding interaction counters by 1/N where N=number of constituent datapoints.

**Details View.** When an aggregate visualization element (e.g. bar) is hovered on, the Details View below the chart shows a table with each datapoint. Lumos captures a mouseover on a table row, treats it as an interaction with the corresponding datapoint, and leaves an interaction trace by updating the table row’s background color (Figure 2) (**DG 2**).

**Attribute Panel.** Like datapoints, Lumos also tracks user interactions with attributes. Each attribute card in the Attribute Panel is colored on a white→blue scale based on the corresponding number of interactions (white=no interaction; darkest blue=most interactions). Lumos captures attribute assignments to encodings (e.g., X, Y) and filters (e.g., Gender=Male) as a proxy for modeling analytic behavior from interactions with attributes. These interactions are totaled and normalized relative to the most interacted attribute to determine the resultant shade of blue, e.g., in Figure 1B, *Genre* has been interacted with most (dark blue) and *Worldwide Gross* has not been interacted with at all (**DG 1**).

#### 3.3.2 Ex-situ Interaction Traces

Lumos allows users to compare their analytic behavior with a target distribution. Attribute cards in the Distribution Panel are colored on a red→grey→green scale (Figure 1G) based on the difference between their respective analytic behavior and underlying distributions (as quantified by Wall et al.’s AD metric [34], Section 2.2) (**DG 4**). In this evaluation, we set the target distribution to the underlying data. A red background indicates that the user’s analytic behavior is different from the target distribution (green background indicates similarity). For example, inspecting the visualization for *Production Budget* shows the analytic behavior peaking around 150M (blue area) when most movies

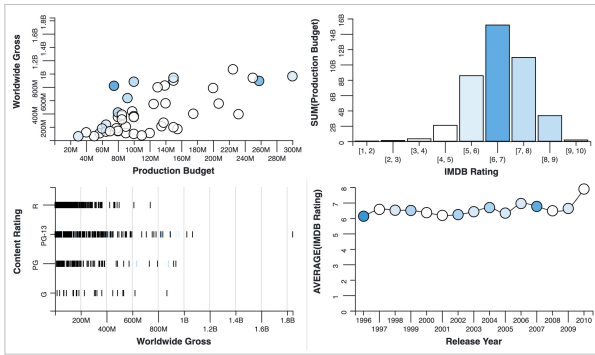


Fig. 3: Lumos captures mouseover interactions on visualization elements (e.g., points, bars) and leaves traces in-situ by coloring (fill) them in shades of blue in proportion to the relative frequency of interactions.

have a budget under 50M (black curve); the magnitude of the deviation is high resulting in a red background. Similarly, the computed analytic behavior (blue bars) on *Content Rating* is more closely matching the underlying data (black strips) resulting in a green background.

### 3.3.3 Configuring Different Target Distributions

Users can define target interactive analytic behavior in multiple ways: (1) by **proportional** interactions across the various attribute distributions of the dataset, (2) by **equal** interactions across the categories of the dataset, and (3) by defining a **custom** target distribution of interactions across the data (DG3). For example, consider a dataset of job applicants, where 50% of applicants identify as male, 40% of applicants identify as female, and 10% of applicants identify as non-binary. A **proportional** baseline would define the target distribution of interactive behavior such that 50% of interactions should be with male applicants, 40% with female applicants, and 10% with non-binary applicants, while an **equal** baseline would set the target distribution of interactions with 33.3% male applicants, 33.3% female applicants, and 33.3% non-binary applicants. If, for instance, diversity is a target in filling this particular role, then a **custom** baseline might be set, where the target interaction distribution is 40% female, 40% non-binary, and 20% male applicants. Figure 4 summarizes these settings in the context of a movies dataset for the Content Rating attribute, and shows (in blue) how the user’s actual analytic behaviors compare to the target. Users can configure **proportional**, **equal**, or **custom** target distributions per attribute in Lumos. In the **custom** mode for categorical attributes, users are presented with an interactive bar chart where they can drag individual bars (each representing a category) to their desired relative weights. For quantitative attributes, users can sketch a target distribution by clicking (to add new quantiles) and dragging points in the presented interactive histogram. While Lumos supports all three target distribution variations, our study (Section 5) utilizes a **proportional** configuration by default. In this evaluation, we focus on obtaining qualitative feedback from users about *awareness of analytic behavior* and defer the study of users defining target distributions to future work.

### 3.3.4 Feedback from Pilot Studies

The aforementioned Lumos interface and interaction design was realized after incorporating feedback from pilot studies with five participants. These helped us fix bugs, add enhancements (e.g., enable resizing the panels), reduce errors and slips (e.g., only showing the Aggregate encoding when applicable), and guide the design choices in the subsequent user study (e.g., we decided to remove the movie *Title* from the dataset as participants reported that their experiential knowledge stimulated saliency biases causing an exploration that neglected the underlying data almost entirely).

## 4 EXAMPLE USAGE SCENARIOS

### 4.1 Scenario 1: Increasing awareness of analytic behavior

Assume Austin is looking for a new home and is exploring a housing dataset in Lumos (Figure 5). After acquainting themselves with

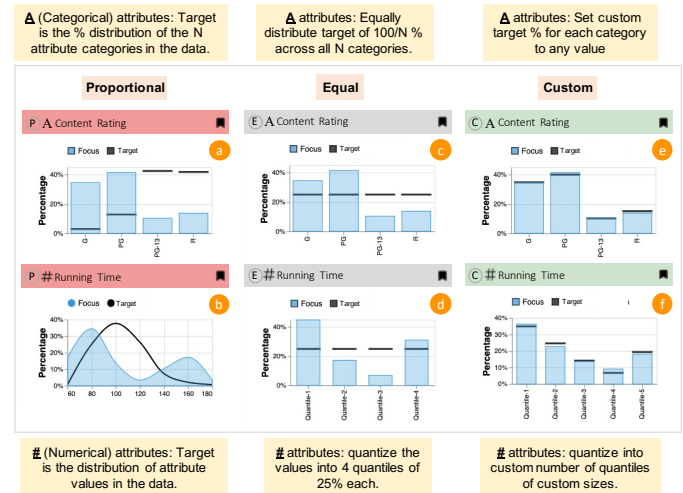


Fig. 4: Lumos supports ex-situ interaction traces for three modes of target distributions (*Proportional*, *Equal*, *Custom*). Lumos presents these targets in the charts as black curves/strips along with user behavior (blue area). Lumos also computes the difference between target and observed behavior and encodes it as the background color of the corresponding attribute card (red, grey, green colors where redder=more different; greener=more similar).

the attributes, they apply three filters that match their criteria:  $\{Home\ Type=Single\ Family; Price \leq \$300K; Satisfaction \geq 7\}$  (Figure 5a). Then, they create different visualizations by specifying encodings (Chart Type, X, Y, and Aggregation) in the Encoding panel (Figure 5b).

While interacting with these different visualizations, they observe visualization elements (e.g., bars, points) changing colors to different shades of blue. For example, in the scatterplot configuration with *Lot Area* (Y axis) and *Year* (X axis) (Figure 5c), Austin observes their focus has been on smaller ( $Lot\ Area \leq 60K$ ) and more recently constructed ( $2009 \leq Year \leq 2010$ ) homes. Similarly, in the barchart configuration with *Foundation Type* (X axis) and *Average(Price)* (Y axis) (Figure 5d), they observe they have not focused on two types of *Foundation Types*:  $\{Brick \ \& \ Tile, Poured\ Concrete\}$  (white).

During their analyses, they also observe different shades of blue in the Attributes Panel (Figure 5e) inferring they have not focused on all attributes equally, e.g., they focused more on *Price* and *Satisfaction* (darker blues), not so much on *Year* and *Foundation Type* (light blues) and not-at-all on *Fireplaces* and *Heating Type* (white).

After acknowledging that they did focus on the blue attributes, they start inspecting the five white attributes. They state they do not care about  $\{Lot\ Config\ and\ Fence\ Type\}$  but regret not focusing on the other three attributes (*Heating Type*, *Fireplaces*, *Central Air*) associated with climate control as the city faces severe winters. Accordingly, they apply new filters and encodings and continue their analyses.

In this way, Lumos helped Austin in house-hunting by making them more aware of their analytic behavior with data and attributes.

### 4.2 Scenario 2: Mitigating biased analytic behavior

Taylor, a loan officer, is using Lumos to analyze loan applications to determine credit-worthiness. After exploring the dataset for a while, they observe a red *Home Ownership Type* attribute card and a green *Age* attribute card (Figure 6a) in the Distribution Panel. They express happiness at not exhibiting any age bias but are concerned that their interactions with *Home Ownership Type* significantly deviate from the underlying data (target) distribution. They click on the card to toggle it open and begin inspecting the visualization. They observe they have unknowingly overemphasized on *Own* and *Rent* and underemphasized on *Mortgage Home Ownership Type*.

Willing to correct their behavior, they apply a (reverse) filter:  $\{Home\ Ownership\ Type=Mortgage\}$  (Figure 6b) and analyze a few previously unconsidered (white) points (Figure 6c). They finally see a greener *Home Ownership Type* card (Figure 6d) and are more content.

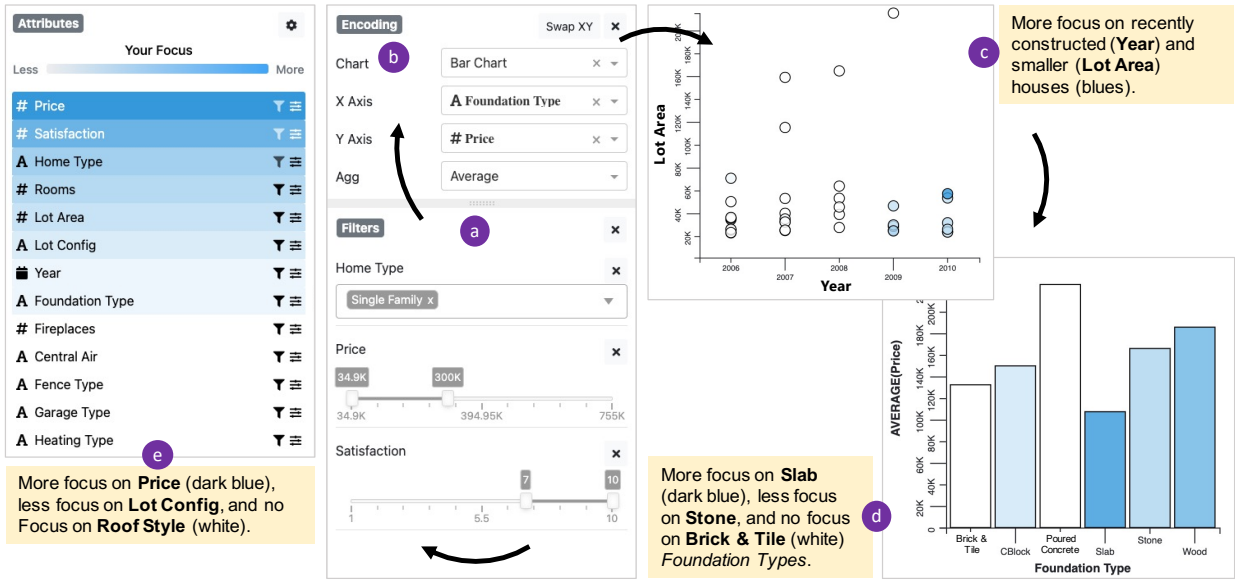


Fig. 5: Scenario 1: Lumos helping Austin be more aware of their data analyses.

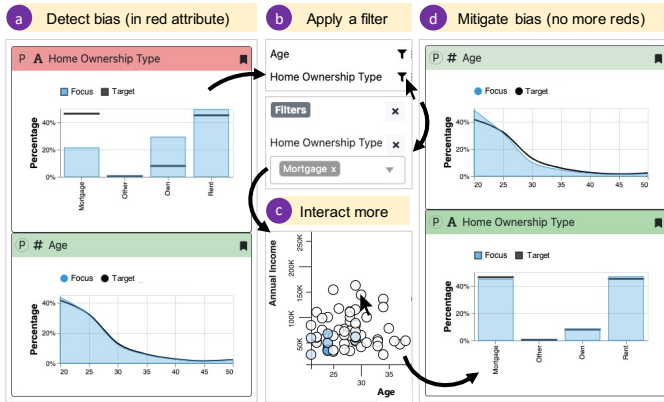


Fig. 6: Scenario 2: Lumos helping Taylor mitigate biased analytic behavior within the *Home Ownership Type* attribute categories.

## 5 EVALUATION

### 5.1 Study Design

We conducted a between-subjects qualitative study with the aim of understanding how Lumos helps users increase awareness of their analytic behaviors. The study protocol was reviewed and approved by our institutional review board.

We recruited 24 participants (Gender: 14 male, 8 female, 1 other, 1 preferred not to say; Age in years:  $\mu=24.04$ ,  $\sigma=3.44$ , median=24, 1 preferred not to say). Participants included students, researchers, and industry professionals with diverse educational degrees (8 bachelors, 7 masters, 9 doctoral) all from a computing or related field (e.g., computer science, human-computer interaction, human-centered computing) with a self-reported visualization literacy  $\geq 3$  (on a 5-point Likert scale). Sessions lasted approximately 1 hour and participants were each compensated with a \$15 Amazon gift card.

Due to the COVID-19 pandemic, we conducted the study remotely using online collaboration tools. The actual study was conducted over the Microsoft Teams (<https://www.microsoft.com/en-us/microsoft-teams/group-chat-software>) teleconferencing software. The experimenter provided participants access to the study environment by sharing their computer’s screen and granting input control.

Participants were randomly divided into either a **Control** or **Awareness** condition, which determined the version of the system they used for the study. Both conditions had roughly equal numbers of par-

ticipants with bachelors, masters, and doctoral degrees (Awareness: bachelors=4, masters=4, doctoral=4; Control: bachelors=4, masters=3, doctoral=5). Participants in the Control condition *did not* see the Distribution Panel (along with the *ex-situ* interaction traces) nor did they see the *in-situ* interaction traces in the Visualization Canvas, Details View, and Attribute Panel. We pre-configured the target distribution to *Proportional* and hid the Settings Panel for both conditions.

Each participant first electronically signed a consent form. Then, depending on the study condition, they saw a 2 minute (Control) or 4 minute (Awareness) video tutorial that demonstrated the features of Lumos. Participants were then asked to perform a practice task using a dataset of cars to interact with and get acquainted with Lumos. After practice, we tasked participants to “analyze a dataset of movies to recommend the characteristics of movies that a movie production company should make next”. The dataset consisted of 709 movies (rows) and 9 attributes (columns): *Production Budget* (#), *Worldwide Gross* (#), *Running Time* (#), *IMDB Rating* (#), *Rotten Tomatoes Rating* (#), *Release Year* (📅), *Content Rating* (A), and *Creative Type* (A). Participants were encouraged to think aloud during the task. Throughout the task, their interactions were recorded. The study ended with a debriefing in which participants discussed their overall experience with the system, provided suggestions for improvements, and completed a questionnaire to rate the usefulness of Lumos features. All sessions were screen- and audio-recorded for qualitative analysis.

### 5.2 Hypotheses

We structure our study analysis according to the following hypotheses:

- H1** Interaction traces will increase awareness of analytic behavior.
- H2** There will be differences in interactive behaviors of Awareness v. Control participants (as measured by bias metrics [34], differences in use of filters, and number of charts created).
- H3** Participants will find the *ex-situ* awareness features to have greater utility than *in-situ* awareness features.
- H4** Participants in the Awareness condition will react to interaction traces in ways to reduce potential biased analytic behaviors.

### 5.3 Results

Below, we present findings from the study and discuss them in context of participant feedback.  $P_A01$ - $P_A12$  and  $P_C13$ - $P_C24$  refer to the 24 participants in Awareness and Control conditions, respectively. Participant quotes and moments of awareness were coded and categorized using affinity diagramming. One administrator came up with an initial set

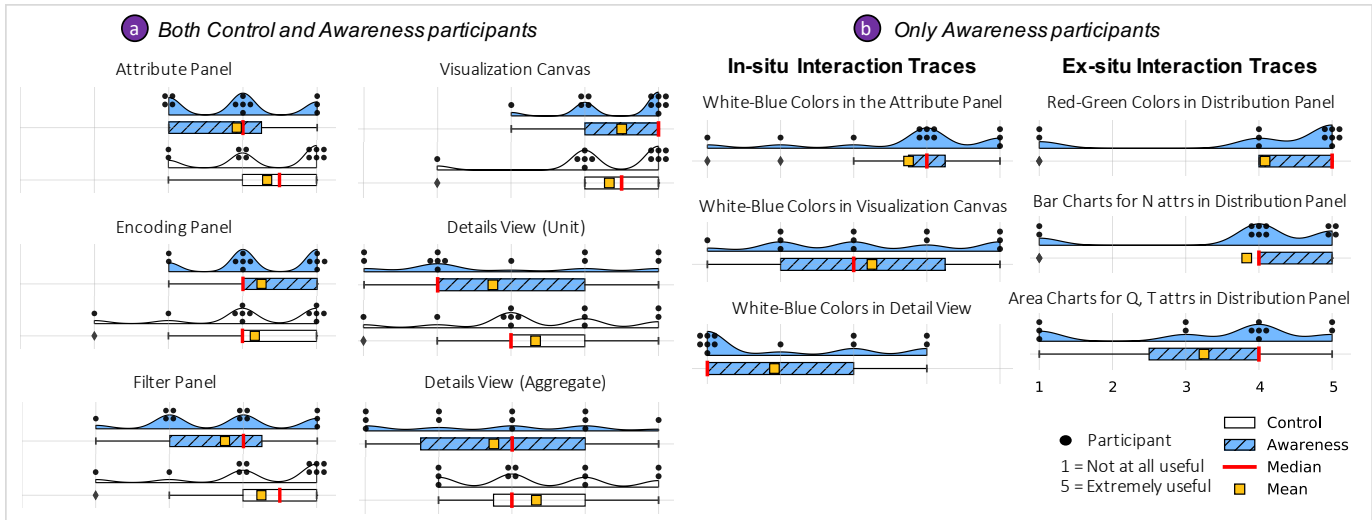


Fig. 7: Summary of usefulness scores of all Lumos features as reported by participants in the post-study questionnaire, as RainCloudPlots [2].

of categories that were then refined during iterations with three other administrators until a consensus was reached.

### 5.3.1 General Feedback

Overall participant feedback was positive with  $P_{A01}$  commenting, “[they] haven’t seen many things like [Lumos] before...really good technique”.  $P_{A09}$  mentioned that “[Lumos] can remove the internal bias of things users think are of the most interest”. Participants found “the ability to keep track of [their] provenance, interaction history [as] interesting” ( $P_{A08}$ ) and “communicating it back [how they are doing] as something [they] would use in [their] tools” ( $P_{A07}$ ).  $P_{A09}$  found the Distribution panel “a great idea to show users what their focus was” and  $P_{A05}$  found it “very helpful as [they] don’t need to create visualizations in the Vis panel for each attribute to see [their] distributions”.  $P_{A05}$  suggested “integrating this tool into existing tools such as Tableau [as] they don’t have a feature that tells [them] what attributes haven’t been used yet” ( $P_{A12}$ ).  $P_{A07}$  suggested “there are lots of use-cases for this technique in journalism and social media, e.g., you have only looked at Trump’s negative tweets, but what about Biden’s?”. Two participants found the Distribution Panel less useful as they “didn’t know exactly what to do about the [red-green] cards” ( $P_{A5}$ ) or felt it “out of focus” on the right side of the screen ( $P_{A10}$ ).

### 5.3.2 Usefulness Scores: Lumos user interface

Figure 7a summarizes Lumos’s usefulness scores ( $1=Not\ useful\ at\ all;$   $5=Extremely\ useful$ ) as reported in the post-study questionnaire:

**Attribute Panel.** Participants generally found the attribute list useful ( $\geq 3$  out of 5,  $median_A=4$ ,  $median_C=4.5$ ) along with “their data types” ( $P_{C21}$ ) “unlike, e.g., Excel where they aren’t always on-screen” ( $P_{C17}$ ).

**Encoding Panel.** 23 participants found the Encoding Panel to be useful ( $median_A=4$ ,  $median_C=4$ , “it is standard in a good way”- $P_{C21}$ ) except  $P_{C23}$  who found it “only slightly useful”. Four participants noted that the system messages to fix incorrect encodings were intuitive and helpful (“it is sometimes hard to know what’s wrong in Tableau”- $P_{A05}$ ) but two found them confusing and suggested the app “prevent [them] from choosing incorrect encodings” ( $P_{C18}$ ) by “filtering out the chart types that are not allowed” ( $P_{A11}$ ). Five participants also suggested additional features (“add color as a third encoding” -  $P_{C24}$ ) and enhancements (“support drag-drop attributes to the Encoding Panel” -  $P_{C17,18,19}$ }, “support text entry for the dropdowns” -  $P_{A12}$ ).

**Filter Panel.** Participants utilized filters ( $median_A=4$ ,  $median_C=4.5$ ) “to remove outliers and to confirm hypotheses about the data” ( $P_{C22}$ ), to see the different values for a categorical attribute ( $P_{C19}$ ,  $P_{C23}$ ), and to mitigate any unconscious biased analytic behavior (e.g., “Comedy and Drama are high percentage in the dataset, and I haven’t interacted with

them at all, so it might be worth my time to look at them.”- $P_{A09}$ ).  $P_{A10}$  did not utilize filters as they were “being more exploratory with [their] analysis and if [they] wanted to look at finer details, [they] would have used more filters.” Three participants also requested enhancements to “specify precise inputs [for Q,T values]” ( $P_{C22}$ ), “allow hover on a value in the categorical filter and highlight in the Visualization” ( $P_{A09}$ ), and “support select- and deselect- all for categorical values” ( $P_{A11}$ ).

**Visualization Canvas.** 23 participants found the Visualization Canvas useful ( $median_A=5$ ,  $median_C=4.5$ ), utilizing it to “observe patterns and outliers” ( $P_{A10}$ ), and “see the different categories and values in the attributes” ( $P_{C14}$ ).  $P_{C18}$  “did not find it useful because a third attribute encoding, e.g., color was not supported”.

**Details View (Unit).** This view received mixed usefulness scores from our participants ( $median_A=2$ ,  $median_C=3$ ).  $P_{C17}$  “liked the Details portion and being able to hover over points for more details.”  $P_{C19}$  and  $P_{A11}$  “didn’t find the Details view for single data points super useful” because they wanted the name of the film to bring prior knowledge to the analysis and spark different hypotheses.<sup>2</sup>

**Details View (Agg).** This view also received mixed usefulness scores from our participants ( $median_A=3$ ,  $median_C=3$ ).  $P_{C18}$  found it to be “really useful because it is not apparent from the bar shape and size that some bars only have one point in them versus some bars having six or seven points”.  $P_{A07}$  said “[they] don’t really hover on things in e.g., a scatterplot but information shown on hovering a bar chart [the details view agg] was awesome because you showed individual data”. Also, “it shows all information in one space.” ( $P_{C20}$ ,  $P_{C24}$ ) and “could be useful for multivariate hypotheses” ( $P_{A10}$ ). Two participants found it “hard to draw conclusions from lists of words and data” ( $P_{C17}$ ,  $P_{C19}$ ) and preferred to see the information visually utilizing the Details View “only as a reference” ( $P_{C19}$ ).  $P_{C18}$  utilized the Details View “because the visualization wasn’t as helpful”.

### 5.3.3 Usefulness Scores: Lumos technique

Participants in the Awareness condition also saw *in-situ* and *ex-situ* (Distribution Panel) interaction traces in the user interface. Figure 7b summarizes the usefulness scores ( $1=Not\ useful\ at\ all;$   $5=Extremely\ useful$ ) as reported in the post-study questionnaire:

**Difference between analytic behavior and target distribution.** Ten participants found the difference between their analytic behavior and the underlying data (red-green coloring in Distribution Panel) very useful ( $median_A=5$ ) “giving a sense if I’m looking at the dataset in an unbiased way” ( $P_{A06}$ ).

<sup>2</sup>Note: the movie *Title* was deliberately not shown in the Detail View to prevent personal experiences with data to influence the analysis.

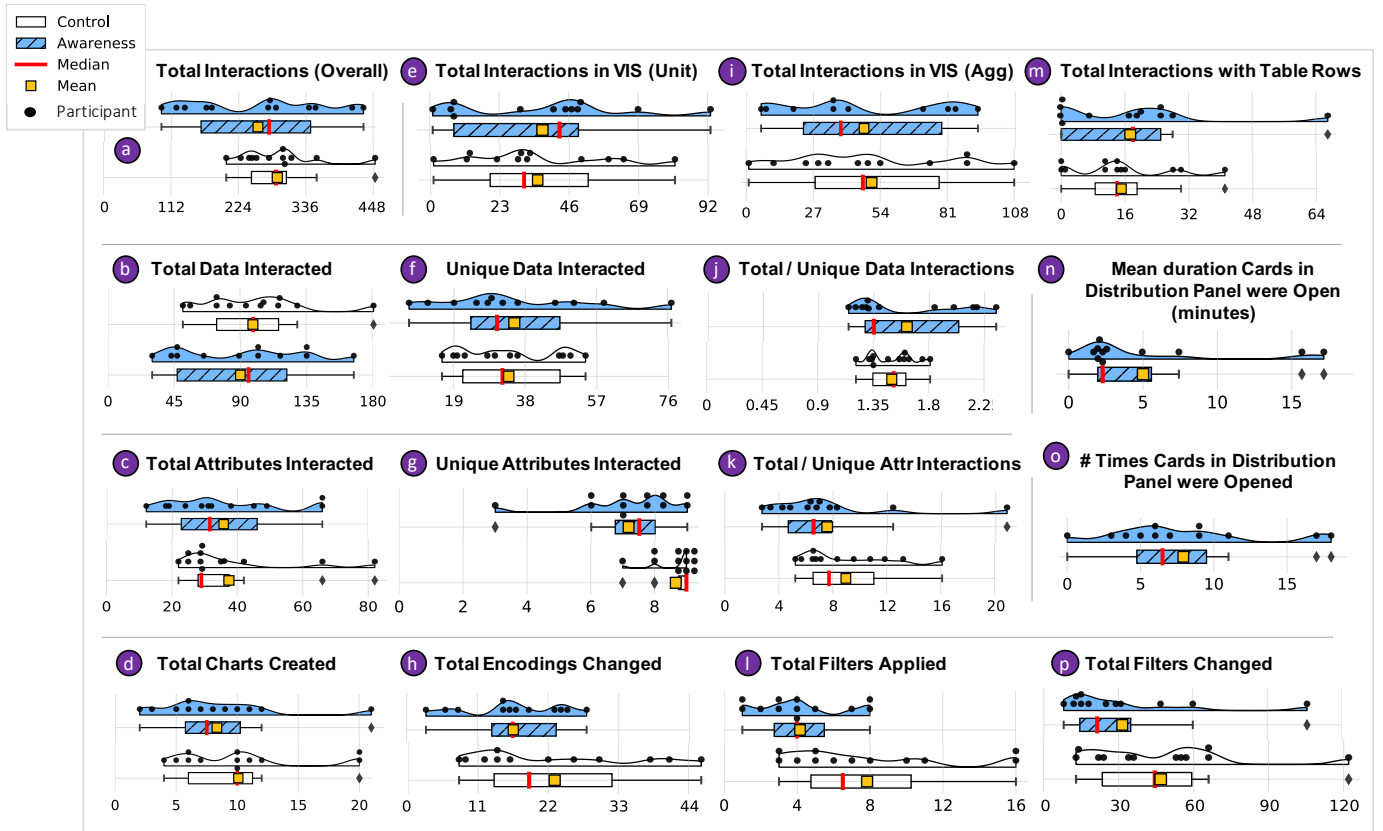


Fig. 8: Summary of quantitative findings across the two experimental conditions, as RainCloudPlots [2]. VIS=Visualization; Attr=Attribute.

**Ex-situ interaction traces in Distribution Panel.** The overall feedback on the *ex-situ* interaction traces in the Distribution Panel was positive; participants found the bar charts for N attributes ( $\text{median}_A=4$ ) more useful than the area charts for Q,T attributes ( $\text{median}_A=4$ ) with  $P_A06$  nicely summarizing, “I can track and channel my focus based on discrete bar charts by applying a filter..but it is difficult to discretize and track a continuous [Q,T] attribute”. For eight users, these real-time traces helped increase awareness (“Geez, I haven’t looked at Drama movies at all”- $P_A07$ ), influencing them to interact differently (e.g., by creating a bar chart with *Genre* to inspect movies of other potentially underemphasized genres) while two participants either ignored them (“I never looked at the individual distributions of attributes”- $P_A12$ ) or preferred to look at them after analysis “as the bars will be moving, and that’s distracting” ( $P_A09$ ).

**In-situ interaction traces in Visualization Canvas.** There was mixed response to the *in-situ* interaction traces (white-blue colors) in the Visualization Canvas ( $\text{median}_A=3$ ). For  $P_A06$ , they “helped in tracking visited points” nudging them to interact with points that were not interacted yet, while for  $P_A05$  they were confusing and distracting, nudging them to re-interact with them.  $P_A08$  did not want the colors to stay persistent but “be able to clear existing interactions and start a new session with a new set of model movies for comparison”.

**In-situ interaction traces in Details View.** Only four participants found the *in-situ* interaction traces (white-blue colors) in the Details View to be useful ( $\text{median}_A=1$ ).

**In-situ interaction traces in Attribute Panel.** Participants generally found the *in-situ* interaction traces (white-blue colors) in the Attribute Panel to be useful ( $\text{median}_A=4$ ). For  $P_A05$ , they helped increase awareness of already-interacted attributes (“I see that I have spent a lot of time on Release Year so I’ll now see something else”) while for  $P_A12$ , they required some time to get acquainted with (“the coloring in the Attributes panel...I didn’t use it initially, and later on it hit me that I had this feature. Once I noticed it, it was very useful”).

**Summary.** Comparing the distributions of scores for the aforementioned features (Figure 7b), participants found the *ex-situ* interaction

traces more useful than the *in-situ* interaction traces, supporting **H3**, consistent with experimental results from [36]. We believe this is because *in-situ* interaction traces are always visible to a user whereas *ex-situ* interaction traces can be used more on-demand without side-tracking the analysis task at hand. Furthermore, *in-situ* traces block an otherwise common attribute encoding channel, *color*, that can be undesirable for and cause inconvenience to some users.

### 5.3.4 Awareness Moments

**In-situ traces in the Attributes Panel.** There were instances when Control participants expressed a need for tracking the already-interacted attributes. For example, we observed  $P_C13$  use hand gestures to recollect and count the attributes that they had already visited and  $P_C14$  exclaimed, “I hope I have interacted with all (attributes)”. Awareness participants, on the other hand, saw the interaction traces and had several instances of *awareness* during their respective analyses. Two participants *acknowledged* their choices (“I don’t think Release Year should matter too much, hence I am not interacting with it.”- $P_A04$ , “I don’t think Running Time is important to me”- $P_A07$ ) while two participants also suggested correcting future course via *interaction* (“I see that I have spent a lot of time on Release Year so I’ll now see something else”- $P_A05$ , “[on seeing a white attribute bar] now I’m going to interact with Running Time”- $P_A22$ ), also supporting **H4**. Two participants reflected upon their choices after the study while they were answering questions pertaining to self-reported focus on individual attributes in the post-study questionnaire (“actually I forgot, had I remembered, it might have been interesting to not click on the same thing over and over.”- $P_A03$ , “I didn’t use the blue attributes panel but now that I see these questions, I would’ve seen them more”- $P_A08$ ). These and the desire for awareness moments by Control participants validate **H1**.

**In-situ traces in the Visualization Canvas and Details View.** Eight participants found the *in-situ* interaction traces in the visualization canvas to be useful; two participants took some time to get acquainted with them (“very useful but I learnt about them slightly afterwards”- $P_A01$ , “I was initially confused but then over use I got used to them

and found them useful in tracking visited points”- $P_A06$ ).  $P_A03$  found the colors to be useful but questioned the technique because “if it is based on [me] hovering on a point again and again, it might not be 100% correct.”  $P_A05$  was distracted and “getting drawn to the visited points (instead of the white un-visited points).” There was minimal commenting on the *in-situ* traces in the Details View but it led to some awareness for  $P_A04$  who hovered on an uninteracted (white) bar in a bar chart and observed “there aren’t many blue rows which means I haven’t been focusing on it.”

**Ex-situ traces in the Distribution Panel.** There were multiple instances of awareness among participants (supporting **H4**).  $P_A04$  verified the interactions traces by comparing it with ground truth (“distribution of my focus on *Running Time* (blue) is representative of the applied filter”).  $P_A05$  reflected upon seeing three red attributes in the Distribution Panel and hypothesized that they were “just thinking aloud and exploring and will (now) follow a more targeted approach”.  $P_A05$  reflected upon seeing a red *Content Rating* attribute (“Seems I didn’t interact with R-rated movies enough so this view nudges me towards those”) and applied a filter to show only R-rated movies.  $P_A07$  reflected upon the white *Drama* category in the Distribution Panel and justified that they “didn’t care about Dramatization movies [...] who cares?” At one point,  $P_A07$  observed many red cards and exclaimed, “this is so biased but whatever.”  $P_A11$  tried to correlate the effects of their interactions with different attributes (“I noticed that *Adventure* is representative of most values in *Rotten Tomatoes Rating* and *IMDB Rating* [...] This is because I mostly interacted with just *Adventure* movies and that caused those attributes [in the Distribution Panel] to be colored green”).  $P_A10$  did not use the Distribution Panel as they “didn’t know how to use it in the context of what [they were] doing”.

### 5.3.5 Interactive Behaviors across Experimental Conditions

Overall, Awareness participants exhibited more diverse behavior in terms of interactions with datapoints and attributes than Control (Figure 8a;  $\mu_A=256$ ,  $\sigma_A=115$ ;  $\mu_C=289$ ,  $\sigma_C=66$ ), validating **H2**.

**Interactions with Datapoints.** Figure 8b shows the distribution of all interactions with datapoints across conditions ( $\mu_A=89.92$ ,  $\sigma_A=43.83$ ,  $\mu_C=98.75$ ;  $\sigma_C=35.67$ ).  $P_A08$ , on seeing the interaction traces, became fearful of skewing their interactions. Both conditions interacted with a similar number of unique datapoints (Figure 8f;  $\mu_A=35.08$ ,  $\sigma_A=20.18$ ,  $\mu_C=33.67$ ;  $\sigma_C=13.36$ ) indicating that the *in-situ* blue colorings did not nudge participants to interact with more white points (in fact,  $P_A05$  felt drawn towards the blue points). Interestingly, this result seems different from the one found by Hindsight [12], where participants tended to interact with more unique points when data were colored based on prior interactions. Figures 8(e,i,m) show the distribution of interactions with (e) datapoints in unit visualizations ( $\mu_A=37.33$ ,  $\sigma_A=27.93$ ,  $\mu_C=35.50$ ;  $\sigma_C=23.68$ ), (i) aggregate visualizations ( $\mu_A=47.3$ ,  $\sigma_A=32.53$ ,  $\mu_C=50.5$ ;  $\sigma_C=33.51$ ), and (m) the Details View table rows ( $\mu_A=17.33$ ,  $\sigma_A=19.11$ ,  $\mu_C=15.08$ ;  $\sigma_C=12.62$ ).

**Interactions with Attributes.** Figure 8(c,d) show the distribution of total ( $\mu_A=35.75$ ,  $\sigma_A=17.8$ ,  $\mu_C=37.42$ ;  $\sigma_C=18.25$ ) and unique interactions ( $\mu_A=7.17$ ,  $\sigma_A=1.64$ ,  $\mu_C=8.67$ ,  $\sigma_C=0.65$ ) with attributes across conditions, respectively. Four awareness participants did not interact with certain attributes as they were not important to their analysis (“I don’t think this [*Running Time*] is important to me”- $P_A11$ ) while two others tried to interact with more attributes to try and mitigate their biased analytic behaviors with attributes ( $P_A05$ ) and datapoints ( $P_A07$ ).

**Charts, Filters, and Encodings.** On average, Control participants created more charts (Figure 8d;  $\mu_A=8.33$ ,  $\sigma_A=5.02$ ,  $\mu_C=10.08$ ,  $\sigma_C=5.28$ ), changed more encodings (Figure 8h;  $\mu_A=16.58$ ,  $\sigma_A=7.86$ ,  $\mu_C=23.00$ ,  $\sigma_C=13.11$ ), and applied (Figure 8l;  $\mu_A=4.17$ ,  $\sigma_A=2.44$ ,  $\mu_C=7.83$ ,  $\sigma_C=4.57$ ) and changed (Figure 8p;  $\mu_A=31.58$ ,  $\sigma_A=28.05$ ,  $\mu_C=46.83$ ,  $\sigma_C=30.50$ ) more filters than the Awareness participants.

**Cards in the Distribution Panel.** Eleven awareness participants utilized the Distribution Panel cards to check the data distribution (black curve) and/or compare their analytic behavior with the underlying data. These participants opened these cards multiple times at different points of their analysis (Figure 8o,  $\mu_A=7.91$  times,  $\sigma_A=5.35$  times) and kept them open for different durations (Figure 8n,  $\mu_A=5.00$  minutes,

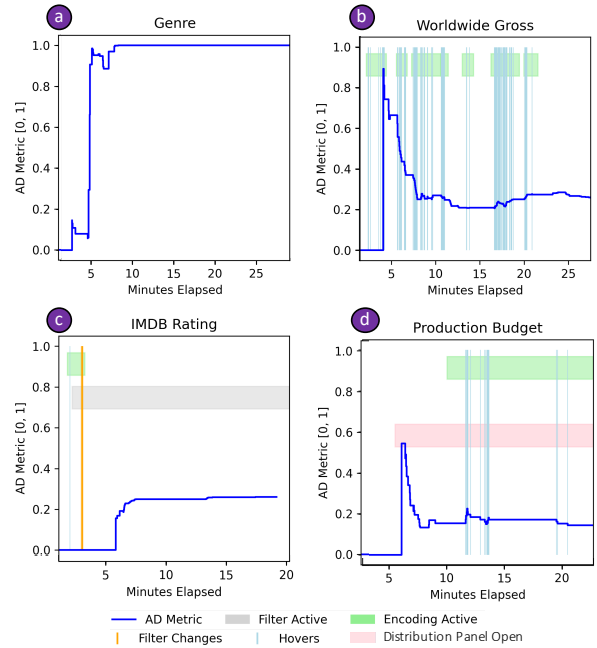


Fig. 9: AD metric values over time for (a)  $P_C16$ , (b)  $P_C13$ , (c)  $P_A09$ , and (d)  $P_A01$  for specific attributes.

$\sigma_A=5.65$  minutes).  $P_A09$  opened the cards but commented that they would prefer analyzing their performance *after* the task was over. One awareness participant did not utilize these cards as they were content with the red-green backgrounds of the closed cards ( $P_A12$ ).

### 5.3.6 Temporal analysis

To understand the cadence of interactions, we plotted hover, filter, encoding, and Distribution Panel interactions over time, and superimposed the AD metric [34] (representing analytic behavior) for each attribute. We found that for some participants who did not interact with these UI components, their analytic behavior was skewed for specific attributes (as evidenced by high AD metric values for *Genre* for  $P_C16$  in Figure 9a). For others, interactions led to significant shifts in analytic behavior. For instance, while  $P_C13$  used *Worldwide Gross* as an encoding in a chart, their analytic behavior toward the attribute became more proportional to the underlying data (Figure 9b). While a filter was applied to IMDB Rating ( $\geq 2.3$  and  $\leq 9.1$ ),  $P_A09$ ’s analytic behavior began to deviate from the underlying data as shown by increasing AD metric value (Figure 9c). Lastly, while we found instances where inspecting the Distribution Panel and accordingly interacting with the visualization appear to reduce biased analytic behavior (e.g., Figure 9d,  $P_A01$ ), such instances were not abundant among our participants. Thus, temporal analysis provides minimal evidence to support **H4**, but interaction traces provided participants opportunities to reflect on their analytic behavior, judging how their process differs from the baseline, and if that difference is appropriate for their task.

## 6 DISCUSSION

### 6.1 Using Color to visualize interaction traces

**Fun, Focus, Distraction, or Inconvenience?** For  $P_A08$ , “interacting [with scatterplot points] and seeing colors change was fun” but for  $P_A05$ , it was a distraction as they were “getting drawn to the visited points [instead of the unvisited points]”.  $P_A09$  noted “highlighting the same data point over and over again skews the distribution visualization, which could be an issue” while  $P_A08$  became “careful to not interact with dots and skew their interaction trace”.  $P_A10$  asserted “a hover does not imply I am interested in that point” and altogether ignored the blue colorings while  $P_A07$  suggested capturing “mouse interactions based on proximity and not an exact hover” to compute analytic behavior. Next, using color to encode interaction traces took



away an encoding channel that could otherwise be used by an attribute. In fact,  $P_A02$  mentioned they “*confused the blue colors with an [attribute] encoding*”.  $P_A12$  also asked if there was “*a reason for having the coloring throughout the visualization?*” Furthermore,  $P_A12$  raised an accessibility concern noting, “*making the tool [Lumos] color-blind safe would be really important*”. Thus, per our participants, using color to visualize interaction traces can be fun, cause a shift in emphasis, cause inconvenience, or be a distraction motivating the need to study alternate encoding approaches.

**If not Color, then what?** While in this evaluation of Lumos, we studied the usage of color (shades of blue, red, and green) to encode interaction traces, there are other visual variables that can be modified to encode and convey the same information, e.g., *stroke color, stroke width, size, shape, orientation, opacity*, etc. It must be noted that some visual variables might work better for certain visualizations than others (e.g., modifying color works better for a scatterplot than a strip plot) and that some variables may not even be applicable for certain visualizations (e.g., it is difficult to modify the shape of a line chart). Exploring this design space will help derive guidelines for effective in-situ and ex-situ visualizations.

## 6.2 The role of target distributions

Target distributions in Lumos serve as a benchmark against which analytic behavior can be compared. For some tasks, meaningful target distributions may exist (e.g., forming committees with specific representation from certain groups). However, it may be harder to articulate a target distribution for other decision-making tasks, in which case standard baselines for comparison are more meaningful. Lumos allows users to modify these or use the data distribution as a default.

## 6.3 False positives in modeling analytic behavior

The Lumos technique of presenting interaction traces can be subject to false positives. For example, users might intentionally not interact with an attribute because it is either not important or they are not interested in it. Labeling these as underemphasized may not be correct, as it was a conscious decision by users to ignore them. Furthermore, a categorical attribute, when encoded along one of the scatterplot axes, can lead to the formation of visual clusters that offload a cognitive task to a perceptual one, rendering that attribute’s filter somewhat redundant. On the other hand, users can also unintentionally neglect aspects of data, e.g., the Attribute Panel may not be able to fit all attributes of a dataset, causing the attributes that are outside the viewport to be potentially neglected during analysis. Lumos helps the user tackle both: by showing unintentionally uninteracted attributes and by allowing users to intentionally set custom target distributions for attributes.

## 6.4 Toward additional mitigation strategies

Based on Lumos results, interaction traces help increase the user’s awareness of their analysis practices, sometimes influencing them to interact and mitigate unconscious biased analytic behaviors. We believe this is a *passive* mitigation strategy since the user has to inspect the difference between their analytic behavior and the target distribution and devise an appropriate strategy, e.g., by applying a filter.  $P_A01$  and  $P_A08$  suggested we implement a more *active* mitigation feature with “*a button to automatically apply a reverse filter [instead of them having to manually apply it]*”, “*especially for continuous attributes*”. For example,  $P_A06$  saw their interactions with different Genres (Concert, Documentary, and Western) and reflected “*[they] should now interact with Drama since that is maximum and these are almost nil*”. They applied a filter to correct their unintended underemphasis but after a few interactions found themselves overemphasizing *towards* Drama movies and reversed the filter. This act of balancing focus across all attributes can lead to frustration, sidelining the analysis task at hand. This is motivation to build mixed-initiative systems that assist the user in mitigating biased analytic behaviors either by acting on-demand (when the user clicks a “Mitigate” button for an attribute) or actively by automatically applying (or removing) a set of filters that negate the overemphasis (or underemphasis).

## 6.5 Lessons Learned

**Encourage users to get lost in their analysis, but use awareness features to remind them.** As described by the guidelines of “fluid interaction” [10], users may become less aware of their own process while performing in-depth analysis. Achieving this level of usability and utility in visualization tools is desirable, but raises the need for awareness functionality such as that in Lumos. Awareness features can help remind users that alternatives should be considered.

**Awareness of one’s own activity is helpful, guidance towards best ways to mitigate may be better.** Participants saw utility in interaction traces toward increasing awareness of their analysis processes. But, what’s next? While users may be aware of potential biased analytic behaviors, there may not be a clear path forward to correct those. Thus it can be fruitful to explore guidance to help users actively mitigate biased analytic behaviors (e.g., by recommending data, visualizations, or filters that may draw a user’s attention to overemphasized or underemphasized parts of the data).

**Different tasks call for different target distributions.** “Biased analytic behaviors” (deviations from a baseline) in context of a movies dataset can likely be chalked up to relatively harmless preferences. However, given different analysis tasks, or different domain contexts (hiring, medical, etc), there may be a much more urgent need to ensure that the target baseline distribution is fit to the task. Lumos can provide the flexibility to specify custom target distributions accordingly.

**Promote awareness while maintaining user agency and control.** While we maintain that user agency and control should be ensured, providing people with awareness of their analysis behavior has merit. At times, these goals may be at odds. For instance, if active strategies are employed for systems to automatically apply bias-mitigating measures, then agency may be compromised. These design decisions should be carefully considered when designing visualizations.

## 7 LIMITATIONS

Lumos currently supports only a small set of visualization types; however, we chose them to test across different aggregation types. Also, analytic behavior is modeled only from interactions, which may not be a complete proxy for attention; in the future, one may consider user gaze or other sources to more accurately approximate it. Lastly, Lumos models analytic behavior by equally weighting the interactions. As  $P_A06$  suggested to “*remove older interactions, say only keep the most recent 100 or 200 of them [as] people lose attention [over time]*”, we may consider interaction recency in the future.

Finally, interactions with aggregate visualizations (e.g., hovering on a bar in a bar chart) are currently considered as  $N$  equally weighted interactions of magnitude  $1/N$  where  $N$  = number of data points belonging to that element. This has variable impact on the metrics due to different statistical tests used to compute the analytic behavior model (AD [34]) depending on the attribute type (e.g.,  $\chi^2$  test for categorical attributes, Kolmogorov-Smirnov test for numerical distributions). Future work can explore alternative computations for models of analytic behavior that may reflect a user’s attention and intentions more precisely.

## 8 CONCLUSION

Lumos is a visual data analysis tool that models analytic behavior of users from their interaction history and provides real-time feedback for awareness and self-reflection of, e.g., overemphasis (or underemphasis) on aspects of data. Our evaluation found that Lumos increases users’ awareness of their analysis behaviors in real-time, promoting reflection upon and acknowledgement of their intentions with the data. These results can have far-reaching implications toward mitigating biased analytic behaviors in decision making contexts, e.g., aid a hiring committee to meet their gender diversity targets and generally foster more transparent analysis processes.

## ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation grant IIS-1813281 and the Siemens FutureMaker Fellowship. We thank the reviewers for their constructive feedback during the review phase. We also thank the Georgia Tech Visualization Lab for their feedback.

## REFERENCES

- [1] J. Alexander, A. Cockburn, S. Fitchett, C. Gutwin, and S. Greenberg. Revisiting read wear: analysis, design, and evaluation of a footprints scrollbar. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1665–1674, 2009.
- [2] M. Allen, D. Poggiali, K. Whitaker, T. R. Marshall, and R. A. Kievit. Raincloud plots: a multi-platform tool for robust data visualization. *Wellcome open research*, 4, 2019.
- [3] J. Angwin, J. Larson, L. Kirchner, and S. Mattu. Machine bias. *ProPublica*, May 2016.
- [4] E. Arroyo, T. Selker, and W. Wei. Usability tool for analysis of web designs using mouse tracks. In *CHI'06 extended abstracts on Human factors in computing systems*, pp. 484–489, 2006.
- [5] S. K. Badam, Z. Zeng, E. Wall, A. Endert, and N. Elmqvist. Supporting team-first visual analytics through group activity representations.
- [6] I. Cho, R. Wesslen, A. Karduni, S. Santhanam, S. Shaikh, and W. Dou. The anchoring effect in decision-making with visual analytics. *IEEE Conference on Visual Analytics Science and Technology (VAST)*, 2017.
- [7] Z. Cutler, K. Gadhave, and A. Lex. Ttrack: A library for provenance-tracking in web-based visualizations. In *2020 IEEE Visualization Conference (VIS)*, pp. 116–120, 2020. doi: 10.1109/VIS47514.2020.00030
- [8] E. Dimara, A. Bezerianos, and P. Dragicevic. The attraction effect in information visualization. *IEEE transactions on visualization and computer graphics*, 23(1):471–480, 2017.
- [9] C. Dunne, N. Henry Riche, B. Lee, R. Metoyer, and G. Robertson. Graphtrail: Analyzing large multivariate, heterogeneous networks while supporting exploration history. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1663–1672, 2012.
- [10] N. Elmqvist, A. V. Moere, H.-C. Jetter, D. Cernea, H. Reiterer, and T. Jankun-Kelly. Fluid interaction for information visualization. *Information Visualization*, 10(4):327–340, 2011.
- [11] M. R. Endsley. Toward a theory of situation awareness in dynamic systems. *Human factors*, 37(1):32–64, 1995.
- [12] M. Feng, C. Deng, E. M. Peck, and L. Harrison. Hindsight: Encouraging exploration through direct encoding of personal interaction history. *IEEE transactions on visualization and computer graphics*, 23(1):351–360, 2016.
- [13] M. Feng, E. Peck, and L. Harrison. Patterns and pace: Quantifying diverse exploration behavior with visualizations on the web. *IEEE transactions on visualization and computer graphics*, 25(1):501–511, 2019.
- [14] G. Gigerenzer. Fast and frugal heuristics: The tools of bounded rationality. *Blackwell handbook of judgment and decision making*, 62:88, 2004.
- [15] D. Gotz, S. Sun, and N. Cao. Adaptive contextualization: Combating bias during high-dimensional visualization and data selection. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*, pp. 85–95, 2016.
- [16] A. G. Greenwald, D. E. McGhee, and J. L. Schwartz. Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology*, 74(6):1464, 1998.
- [17] C. Gutwin. Traces: Visualizing the immediate past to support group interaction. In *Graphics interface*, pp. 43–50. Citeseer, 2002.
- [18] J. Heer, J. Mackinlay, C. Stolte, and M. Agrawala. Graphical histories for visualization: Supporting analysis, communication, and evaluation. *IEEE transactions on visualization and computer graphics*, 14(6):1189–1196, 2008.
- [19] W. C. Hill, J. D. Hollan, D. Wroblewski, and T. McCandless. Edit wear and read wear. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 3–9, 1992.
- [20] C. E. Jordan, S. J. Clark, and M. A. Waldron. Gender bias and compensation in the executive suite of the fortune 100. *Journal of Organizational Culture, Communications and Conflict*, 11(1):19, 2007.
- [21] Z. Liu and J. Heer. The effects of interactive latency on exploratory visual analysis. *IEEE transactions on visualization and computer graphics*, 20(12):2122–2131, 2014.
- [22] G. A. Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, 101(2):343, 1994.
- [23] J. Nielsen and K. Pernice. *Eyetracking web usability*. New Riders, 2010.
- [24] C. North, R. Chang, A. Endert, W. Dou, R. May, B. Pike, and G. Fink. Analytic provenance: process+ interaction+ insight. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pp. 33–36, 2011.
- [25] A. Ottley, R. Garnett, and R. Wan. Follow the clicks: Learning and anticipating mouse interactions during exploratory data analysis. In *Computer Graphics Forum*, vol. 38, pp. 41–52. Wiley Online Library, 2019.
- [26] E. D. Ragan, A. Endert, J. Sanyal, and J. Chen. Characterizing provenance in visualization and data analysis: an organizational framework of provenance types and purposes. *IEEE transactions on visualization and computer graphics*, 22(1):31–40, 2015.
- [27] A. Sarvghad and M. Tory. Exploiting analysis history to support collaborative data analysis. In *Proceedings of the 41st Graphics Interface Conference*, pp. 123–130, 2015.
- [28] A. Sarvghad, M. Tory, and N. Mahyar. Visualizing dimension coverage to support exploratory analysis. *IEEE transactions on visualization and computer graphics*, 23(1):21–30, 2016.
- [29] R. Sauerborn, P. Berman, and A. Nougara. Age bias, but no gender bias, in the intrahousehold resource allocation for health care in rural burkina faso. *Health transition review*, pp. 131–145, 1996.
- [30] Y. B. Shrinivasan and J. J. van Wijk. Supporting the analytical reasoning process in information visualization. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1237–1246, 2008.
- [31] A. Skopik and C. Gutwin. Improving revisitation in fisheye views with visit wear. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 771–780, 2005.
- [32] A. C. Valdez, M. Ziefle, and M. Sedlmair. Priming and anchoring effects in visualization. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):584–594, 2018.
- [33] E. Wall, A. Arcalgud, K. Gupta, and A. Jo. A markov model of users' interactive behavior in scatterplots. In *2019 IEEE Visualization Conference (VIS)*, pp. 81–85. IEEE, 2019.
- [34] E. Wall, L. M. Blaha, L. Franklin, and A. Endert. Warning, bias may occur: A proposed approach to detecting cognitive bias in interactive visual analytics. In *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 104–115. IEEE, 2017.
- [35] E. Wall, L. M. Blaha, C. Paul, and A. Endert. A formative study of interactive bias metrics in visual analytics using anchoring bias. *Proceedings of the 17th IFIP TC 13 International Conference on Human-Computer Interaction (INTERACT'19)*, 2019.
- [36] E. Wall, A. Narechania, A. Coscia, J. Paden, and A. Endert. Left, right, and gender: Exploring interaction traces to mitigate human biases. *IEEE Transactions on Visualization and Computer Graphics*, 2021. To appear.
- [37] W. Willett, J. Heer, and M. Agrawala. Scented widgets: Improving navigation cues with embedded visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1129–1136, 2007.
- [38] Z. Zhou, X. Wen, Y. Wang, and D. Gotz. *Modeling and Leveraging Analytic Focus During Exploratory Visual Analysis*. Association for Computing Machinery, New York, NY, USA, 2021.