EuroVis Workshop on Trustworthy Visualization (TrustVis) (2019)
R. Kosara, K. Lawonn, L. Linsen, and N. Smit (Editors)

Detection of Confirmation and Distinction Biases in Visual Analytics Systems

. A. Nalcaci, D. Girgin, S. Balki, F. Talay, H. A. Boz and S. Balcisoy

[1]Sabanci University, Faculty of Engineering & Natural Sciences, Istanbul, Turkey

Abstract

Cognitive bias is a systematic error that introduces drifts and distortions in the human judgment in terms of visual decomposition in the direction of the dominant instance. It has a significant role in decision-making process by means of evaluation of data visualizations. This paper elaborates on the experimental depiction of two cognitive bias types, namely Distinction Bias and Confirmation Bias, through the examination of cognate visual experimentations. The main goal of this implementation is to indicate the existence of cognitive bias in visual analytics systems through the adjustment of data visualization and crowdsourcing in terms of confirmation and distinction biases. Two distinct surveys that include biased and unbiased data visualizations which are related to a given data set were established in order to detect and measure the level of existence of introduced bias types. Practice of crowdsourcing which is provided by Amazon Mechanical Turk have been used for experimentation purposes through prepared surveys. Results statistically indicate that both distinction and confirmation biases has substantial effect and prominent significance on decision-making process.

CCS Concepts

• ***Human-centered computing*** → ***Empirical studies in visualization; Visualization design and evaluation methods;***

1. Introduction

Visual perception is the combination of physical and thoughtful design that provides the ability to interpret the environment and the information in a way that it creates meaning. Notwithstanding, visual perception of a visual analytic system is prone to certain deceptions, caused by various biased environments akin to visual distortions such as data exposition or symbolic indiscretion.

According to Nussbaumer et al. [NVH*16], Cognitive bias is described as a systematic error with regard to statistical and sequential process management resultantly to the representation of uncertain, complex and/or faulty information. The presence of cognitive bias outgrows drifts and distortions in the human judgment, causing an excessive selection and tendency towards the dominant instance, thus having a substantial position in the evaluation of data visualizations. As Dimara et al. [DBD16] also points out, cognitive bias has a significant effect on users by means of incorrect decisions and inefficiencies in visualization-supported analytic processes that require visual perceptions. On the other hand, Akl et al. [AT16] defines the presence of cognitive bias as an ineluctable influence over the perception of humans.

To exemplify, Confirmation Bias is described as the human tendency to search for, collect, interpret, analyze or recall information in a way that confirms the operator's initial decision and preferences. As Jorgensen [JP15] describes, confirmation bias functions under the principle "First impression determines the action". According to existing studies, usual heuristic errors involve confirmation bias, which characterizes people's approaches to receive the confirmatory corroboration of a pre-existing hypothesis and dismiss contrary information.

Furthermore, Distinction Bias is another cognitive bias type that is described, by Christopher et al. [ACS*17], as the "misprediction and mischoice due to joint evaluation". Moreover, Christopher et al. [ACS*17] emphasizes the utility function of an attribute may differ between single evaluation and joint evaluation. When people in joint evaluation make predictions or decisions about events in single evaluation, they rely on to their joint evaluation preferences instead of their single evaluation preferences and overpredict the Distinction that different values of an attribute will make to their happiness in single evaluation. This overprediction is mentioned as the Distinction Bias (Christopher et al. [ACS*17]).

Additionally, Cho et al. [CWK*17] states the significance and influence of Anchoring Effect as the tendency of focusing a singular element immoderately and ignoring any further information, by indicating its regulation over visual and numerical anchors through a systematic study and analysis in order to point out the fact that the effect causes an inevitable biased decision selection and statistical outcomes.

Another example that can be pointed out is the distinctive influences of word length, character height and word width over the human assessment mechanism through the implementation of data

DOI: 10.2312/trvis.20191185

encoded font size, according to the examinations of Alexander et al. [ACS*17]. The study shows that such factors in a visual analytics system results in having negative impacts on decision-making process when the utilization of comparative font sizes are appertained in opposite directions.

The main goal of this study is to disclose the fragility of decision-making process through the implementation of criminal-based data visualizations in terms of distinction and confirmation biases. The implementation of two different data visualizations with the same data set has been carried out with the application of crowdsourcing by the publication of an online survey for both bias types

2. Objectives

In overall, main objective is to show that decision making processes are prone to manipulation and can be misled towards a particular outcome with respect to the fashion of data visualization and contents. For this purpose, an attentive work flow of the research has been planned and carried out by means of detecting the existence and effects of predetermined cognitive bias types that are being confirmation and distinction biases.

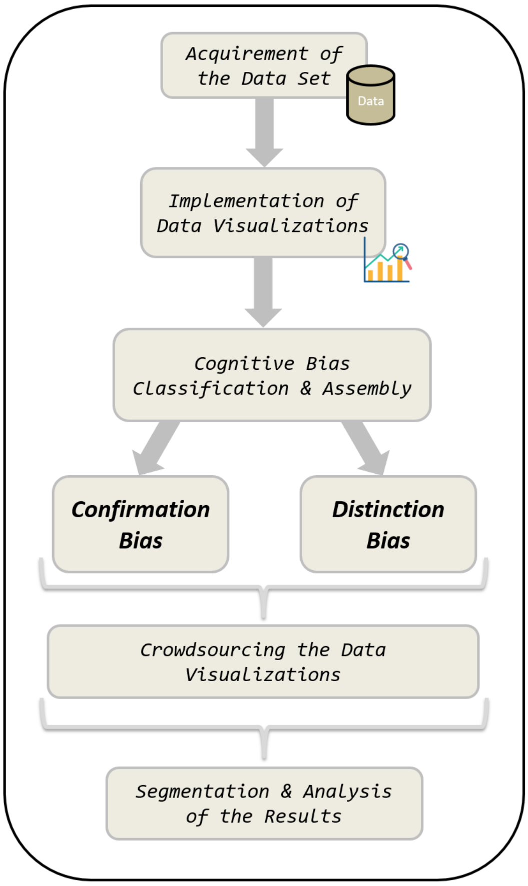

Figure 1: *General overview of the workflow and experimentation process.*

Figure 1 (above) depicts the overall research activities and software components that are acquired and utilized through the course of this study with the addition of illustrating the objective distribution over a six-month interval.

The data visualizations have been carried out distinctly with respect to the two bias types and dispatched in the form of two surveys in which users answered both biased and unbiased displays of the same data. Surveys have been conducted through crowdsourcing by the use of Amazon Mechanical Turk and SurveyMonkey. Subsequently, responses have been analyzed along the lines of legitimate hypothesis testing methods to evaluate the disparity between biased and unbiased versions (detailed information is provided in Section 4).

3. Related Work

Nowadays, cognitive bias research is a prevalent and challenging area in visual analytics systems. Nonetheless, it is not totally obvious at what point it can be applicable in visual analytics.

Since it is a rather difficult task to implement such cognitive bias detection mechanism into laboratory studies, Pohl et al. [PWP*14] proposes a different way for mitigating the bias which is providing context and activate background knowledge. In this study, we also provide a context and background knowledge while creating our detection method along the lines of confirmation and distinction biases.

Moreover, different perspectives have been studied in order to create a common language in research and improvement efforts by promising researchers to select the perspectives planned in their work. Correspondingly, Wall et al. [WBPC18] explains four different viewpoints on human bias that are mainly in relation to visual analytics. She continues by discussing the effect of bias on users by means of incorrect decisions and inefficiencies in visualization-supported analytic processes.

Though there are various studies which retains the introduction priming and anchorage of biases in scatter plots affects the task of class separation, the effects of cognitive bias in a visual analytical system is yet to be explored.

4. Experimental Setup

4.1. Preparation

The initial phase was to gather a suitable data set by means of detecting cognitive bias in decision-making process. For that purpose, a data set that includes the cumulative crime rates in various US cities from 1975 to 2015 has been unearthed from Kaggle Inc. and organized accordingly.

For simplicity purposes, the data of four different cities (cities that create the major distinction in the data set) with the time span from 1995 to 2015 (for distinction bias only) have been extracted and visualized. Thereupon, bias types have been determined and examined one by one by means of the illustration of distinct types in an affiliated context. Following that, pre-obtained data set has been visualized in a bias-detective technique and experimented through crowdsourcing in the course of discerning the influence level of cognitive bias in visual analytics systems.

4.2. Bias Modelling

During the preparation process distinction and confirmation biases have been investigated and visually implemented for the detection of cognitive bias in decision-making process. A data set that is related to the cumulative crime rate in numerous US cities and states have been gathered and visualized in virtue of detecting confirmation and distinction biases in visual analytics systems.

Implementations of bias types were distinct and implemented with reference to cognitive bias exposure. Particularly, confirmation bias affiliated visualizations utilized for the purpose of detecting individuals approaches to receive the confirmatory corroboration of a pre-existing hypothesis and dismiss contrary information. Following that, distinction bias affiliated visualizations utilized in a distinctive technique by representing the data set in equivalently distributed pairs that corresponds to respective data visuals.

Ultimately, in accordance with the results of questionnaires, coinciding hypothesis tests and statistical observations have been implemented to detect whether participants made biased decisions or not.

4.2.1. Confirmation Bias

Confirmation Bias is the human tendency to search for, collect, interpret, analyze or recall information in a way that confirms the operator's initial decision and preferences. As Jorgensen et al. [JP15] points out, it functions with the base principle of "First impression determines the action".

To measure the effects of this bias type, two different visualizations have been created in order to fully grasp the misguided expectations of an information system which is caused by confirmation bias induced distorted user acceptance. Initially, a bar chart has been generated and provided to the participants that displays the crime rates of four different districts of the USA. At the same time, a node graph based visualization have been carried out in order to detect the distinctive and distortive effect of confirmation bias.

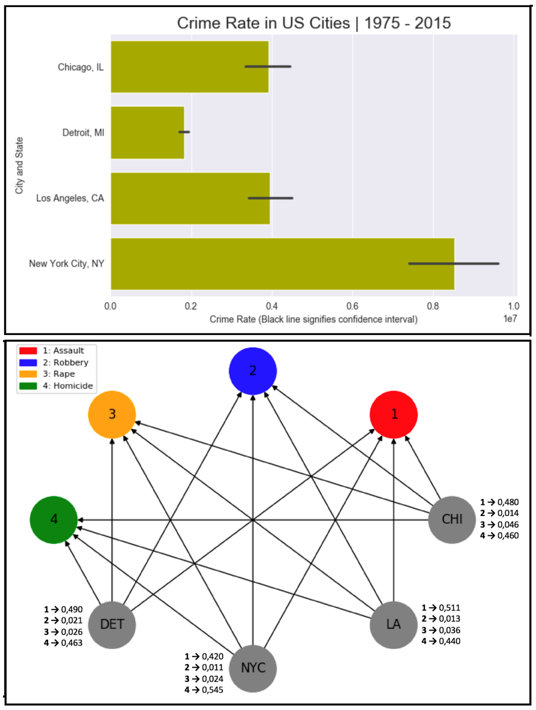

Figure 2: *Confirmation bias affiliated visualizations. Cumulative crime rate of 4 different cities is provided with their confidence interval in the form of a histogram. Rates of distinct crime types is provided in the form of a nodule graph, each with their appropriate crime rate values for distinct types. The question of 'Which of the US cities has the highest combined assault and robbery?' is asked to the participants.*

Depicted in Figure 2 is the Confirmation Bias affiliated visualizations represented in the form of a histogram and node graphs in order to detect the user's adherence to the confirmatory data representation. The same survey question is provided for both charts; Which of the US cities has the highest combined assault and robbery?. Primary expectation of this part of experimentation was to receive a generic propensity towards the option which has the highest cumulative crime rate (which is New York City).

4.2.2. Distinction Bias

Distinction Bias is described as a cognitive bias type that elaborates on the tendency of decision-making in a distinctive fashion when one or more data sets represented simultaneously compared to a seperate provision.

For the purpose of observing the effects of the distinction bias, two different methods of data representation techniques have been used to illustrate the total crimes in US cities between 1995 and 2015. In the first survey, total crimes for Chicago, Detroit, Los Angles and New York are shown separately. In the second survey, total crimes for all cities are plotted on the same graph in order to measure the effect of the distinction bias since the aim is to demonstrate the same data with different visualizations to detect the distinction bias. Affiliated graphs can be observed in Figure 3.

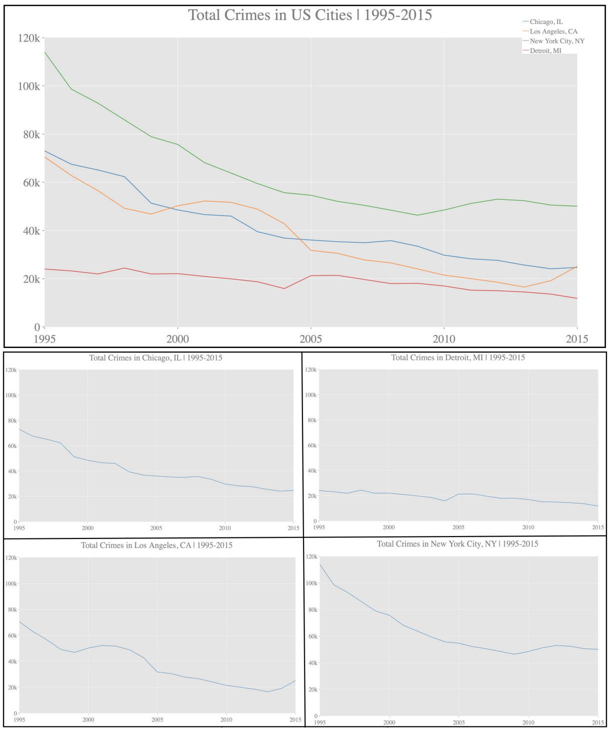

Figure 3: *Distinction bias affiliated visualizations. Crime rates of 4 different US cities have been provided in an aggregate form (unbiased) in the first survey and represented distinctively (biased) in the second survey with the question of 'Which of the US cities has the highest cumulative crime?'.*

The question of 'Which of the US cities has the highest cumulative crime?' is provided to the participants to detect whether if participants will evaluate through a different approach when the data is provided distinctively compared to being provided simultaneously. Even though there was no such expectation of receiving totally dissimilar results and demographics between two surveys, apprehensions was that there will be a noticeable tendency towards the frequency of selection of a second alternative (and there was) during the experimentation that will conclusively indicate the influence level of distinction bias on decision-making process in excess of visual analytics systems and delineations.

5. Results & Conclusion

Both surveys have been published via Amazon Mechanical Turk. A total of 200 people (100 for each) have taken the survey with 93% average completion rate and 243 total responses. Figure 4 depicts the overall statistics for both surveys. In general, we observed that cognitive bias had an absolute effect on participant's decision-making through the analysis of the frequency of selections.

Even though experimentation results indicate a certain level of bias existence through visual observation, implementation of chi-squared tests to the distinction bias affiliated results have been carried out with the purpose of depicting the discrepancy between the

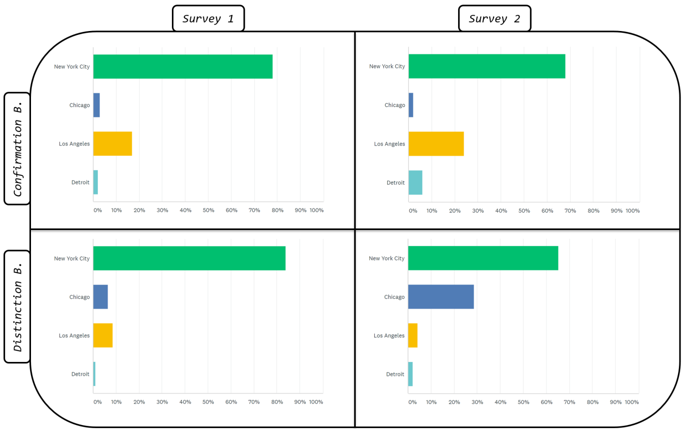

Figure 4: *Demographics of experimentation results. A statistical validation of the existence of distinction based bias can be observed from the hypothesis testing of distinction bias affiliated question results and a clear tendency towards the incorrect answer (New York City) can be observed for confirmation bias affiliated question results in both surveys.*

mean values of question variants. Upon hypothesis testing, we observed that the p-value of the selection set is less than the significance level of the problem set. In particular, test results indicated that the p-value being approximately 8.58e-25 while significance level being 0.05, thus substantiating the essential existence of distinction bias in decision-making.

Apart from Distinction Bias, congregated results immensely indicate the subsistence of confirmation bias in participant's decision-making due to the fact that the highest rate of selection being New York City whereas the true answer for the question was Los Angeles. In general, most participants made a biased selection by adhering to the preexisting cumulative visual which depicts the New York City being first in cumulative crime rate. However, the node representation of the same data clearly indicates the correct answer even though it contradicts the preexisting information in a visual fashion. Nonetheless, experiment results suggest that participants went for their preexisting preference owing to New York City having the highest average frequency of selection with 73% while the true answer, Los Angeles, having the average of 20.5

Ultimately, through statistical analysis and hypothesis tests, the experiment results indicate the misdirecting influence of confirmation and distinction biases on decision-making through data visualizations. By the feasible examination of test results and visual observations, we have failed to reject the hypothesis of the existence of confirmation and distinction biases in decision-making process.

References

[ACS*17] ALEXANDER E., CHANG C.-C., SHIMABUKURO M., FRANCONERI S., COLLINS C., GLEICHER M.: Designing interventions to mitigate cognitive biases in human decisions. *IEEE Transactions on Visualization and Computer Graphics 24*, 8 (July 2017), 2397–2410. doi:10.1109/TVCG.2017.2723397. 1, 2

[AT16] AKT N., TEWFIK A.: Designing interventions to mitigate cognitive biases in human decisions. *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)* (Sept. 2016). doi:10.1109/MLSP.2016.7738838. 1

[CWK*17] CHO I., WESSLEN R., KARDUNI A., SANTHANAM S., SHAIKH S., DOU W.: The anchoring effect in decision-making with visual analytics. *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)* (Oct. 2017). doi:10.1109/VAST.2017.8585665. 1

[DBD16] DIMARA E., BEZERIANOS A., DRAGICEVIC P.: The attraction effect in information visualization. *IEEE Transactions on Visualization and Computer Graphics 23*, 1 (Aug. 2016). doi:10.1109/TVCG.2016.2598594. 1

[JP15] JORGENSEN M., PAPATHEOCHAROUS E.: Believing is seeing: Confirmation bias studies in software engineering. *2015 41st Euromicro Conference on Software Engineering and Advanced Applications* (Aug. 2015). doi:10.1109/SEAA.2015.56. 1, 3

[NVH*16] NUSSBAUMER A., VERBERT K., HILLEMANN E.-C., BEDEK M. A., ALBERT D.: A framework for cognitive bias detection and feedback in a visual analytics environment. *2016 European Intelligence and Security Informatics Conference (EISIC)* (Aug. 2016). doi:10.1109/EISIC.2016.038. 1

[PWP*14] POHL M., WINTER L. C., PALLARIS C., ATTFIELD S., WONG B. L. W.: Sensemaking and cognitive bias mitigation in visual analytics. *IEEE Joint Intelligence and Security Informatics Conference* (2014). doi:10.1109/jisic.2014.68. 2

[WBPC18] WALL E., BLAHA L., PAUL C. L., COOK K.: Four perspectives on human bias in visual analytics. *Journal of Personality and Social Psychology* (2018), 29–42. doi:10.1007/978-3-319-95831-6_3. 2