Examining Whether Approximate Visualizations Impact Perception of Corrected Visualizations

Dominik Moritz and Daniel A. Epstein

Abstract—Data analysts often have to work with an approximate model before more precise results are available. However, after they have seen the accurate results, are they still biased by the approximation? In this paper we present a controlled experiment with 34 participants on Mechanical Turk to investigate whether exposure to approximate results biases user's recollection of the precise data. In our study, we aimed to simulate an exploratory data analysis session with an approximate query processing (AQP) system, a common class of systems used to accelerate data exploration. Our goal was to understand whether analysts are biased by approximate data and whether highlighting the difference reduces this bias. We found that analyst's recollection of precise data is biased but have no conclusive results about whether highlighting differences reduces the bias. The results from this paper are relevant for AQP systems and count extend to other applications in journalism and science.

Index Terms—Approximate visualizations, perception

1 Introduction

As data science grows in interest and importance, data analysts want to derive insights from increasingly large datasets. Depending on the dataset, it could take a system seconds or minutes to construct accurate summary visualizations of the data. The time to generate these visualizations drastically impedes the process of exploratory data visualization [1], where analysts often consider many visualizations in sequence. A common technique to enable exploratory analysis of large datasets is to sample the data for immediate consumption [2]. Precise results considering the entire dataset are either updated progressively [3] or asynchronously [2], [4].

By their nature, the immediate visualizations are not precise. Errors can affect the observations analysts make and thus the conclusions they derive from the data. This work considers the perceptual question of whether temporary exposure to imprecise visualizations in exploratory data analysis influences people's recognition of the data. We further explore whether the design of the precise visualization can reduce the impact of the temporary exposure to imprecise data.

Specifically, this paper seeks to answer the following questions:

1) Does seeing an imprecise visualization generated from sampled data prior to a precise visualization impact people's recollection of the data viewed?
2) Is this impact mitigated by highlighting the change from the imprecise visualization in the precise visualization?

To answer these questions, we conduct a controlled experiment simulating an exploratory data analysis process with a dataset of flights in the United States. Participants first viewed two approximate visualizations generated by sampling the flight data, then viewed the same visualizations updated with the precise data. Our experiment evaluated varying levels of error and different methods for highlighting the change in the precise visualizations.

Through an 38-person experiment on Amazon Mechanical Turk, we found that exposure to an incorrect visualization biases the recollection of data from the precise visualization. However, we did not see enough evidence to support the hypothesis that highlighting the differences reduces this bias.

2 Related Work

This paper builds from literature on exploratory data analysis visualizing uncertainty, and studies of human perception of visualizations.

2.1 Exploratory data analysis for large datasets

In exploratory data analysis (EDA) [5] analysts example multi-dimensional data by looking at the distributions and correlations of fields. This process is iterative and an analyst might look at dozens of graphs as they get to know the data [6]. Some of these charts contain observations that allows analysts to move on to the next step of their analysis [7]. For instance, when examining a dataset of flights, an observation might be "Most flights are out of New York State."

A Visualization system must be fast enough to enable fast iterations; with delays of more than 500ms analysts become less effective [1] or even lose their flow of thought if delays are more than a second [8].

To enable fast response times over large data, visualization system use sampling and approximation to reduce the amount of data that has to be evaluated to compute the results of a query [9], [10], [11].

The HCI community found that approximate visualizations can be used in exploratory data analysis [2], [3]. However, there is no prior work on how exposure to approximate

• *Dominik Moritz and Daniel Epstein are with the University of Washington.*

Manuscript received November 14, 2016.

visualizations biases the observations and insights analysts make.

2.2 Visualizing Uncertainty and Human Perception of Visualizations

There are a number of techniques for communicating uncertainty using visualizations [4], [12]. However, often users often struggle to correctly interpret the uncertainty or even draw incorrect conclusions [13]. Moreover, in many models in the real world no statistical model is available to estimate the uncertainty so that the approximate visualizations in our study do not show uncertainties.

Borkin et.al. [14] and others have studied the memorability of data visualizations. Prior work distinguishes memorability of the visualization itself and the underlying data. In this paper, we only focus on what data users remember because the mental models users build of the data are what influences their decisions on what to look at next. To the best of our knowledge there are no studies on the biases that exposure to imprecise visualizations introduces.

3 Experiment Description

3.1 Research Questions

The questions we want to answer two questions about participants recollection of the data form a precise data when they see an approximate, incorrect visualization first.

RQ1 : Do participants bias their recollection towards the data from the approximate visualization?

RQ2 : Is this bias reduced by showing participants the approximate and precise data or a visualization of the difference as opposed to only the precise data?

We designed our study to answer these two questions in a controlled experiment that mimics a real data analysis session. RQ2 of course only makes sense if the answer to RQ1 is yes but we have reason to believe it is.

3.2 Experimental Design

We sought to create an experiment which mimicked the experience of Exploratory Data Analysis with an approximate query system while still maintaining control over participant experiences. In approximate query systems, users often internalize approximate visualizations, using the gained knowledge to inform the next visualization to create. Systems, such as Pangloss [2], can later generate visualizations of the precise results. However, the user may have moved on to investigate a new visualization before the precise result is computed.

The study is online at https://domoritz.github.io/bias-study/index.html.

3.2.1 Study Flow

Our study flow (Figure 1) aimed to mimic the experience of a data scientist analyzing a large dataset about flights in an approximate query processing system. Participants viewed two approximate visualizations followed by two precise visualizations and then asked questions about the data in them. Before we showed the first visualization, we instructed each participant to focus on a particular airline and state. This mimics a data analysis session where analysts form hypotheses that affect what they focus on rather than trying to look at every possible aspect (e.g. airlines, states) with equal attention. The study enforced participants view the visualizations for at least 10 seconds before advancing. To ensure that participants had considered the approximate data, we asked a few questions after viewing each approximate visualization informed by Brehmer and Munzner's work [15]. Answering these questions simulated how a data analyst would consider and reflect on their data. The questions also provided us a baseline level of comprehension of visualizations of our participants. After viewing both precise visualizations, participants answered a larger set of questions similar to the ones we asked about the visualization of the approximate data. Participants were asked to rate their confidence in their answer to each question in both sets on a sliding scale from "I don't remember" to "I'm certain". Participants were informed that the first two visualizations showed approximations of their data, allowing us to mimic the experience of looking at an approximate query result.

3.2.2 Visualizations in the Study

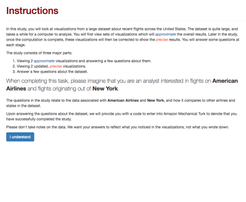
The visualizations participants saw were based on the same dataset about flights in the United States [1]. All visualizations were bar graphs, which are common in approximate query systems [2]. The y-axis varied between the two visualizations, showing either the state the flight left from or the airline the flight was operated by. The length of the bar encodes the number of flights operated by a particular airline or the number of flights out of different states. The bars were sorted in descending order so that the state or airline with the most flights is at the top. We chose horizontal bar charts so that long labels are easier to read. To visually separate visualizations of approximate and precise data, we colored the bars approximate visualizations orange (Figure 2a) and precise visualizations blue (Figure 2b).

We used one month of flight data to generate the visualizations. The precise visualization of number of flights per state uses data where the airline was "JetBlue" and the visualization of the flight per airline uses data from "New York State". The approximate visualizations use a random subset of the data for the precise visualizations with replacement. We then computed the approximate values by dividing the count values by the fraction of rows used in the sample.

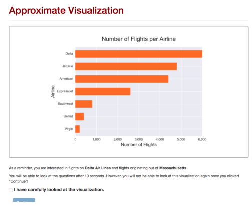
In the visualizations of approximate data, the values can be different because the data is based on a sample. Groups can be missing entirely if there was no example of a particular state or airline in the random sample (e.g. Figure 2a). Moreover, the order of bars can be incorrect because we always sort by the values. The questions we ask participants of our study are designed to fit these errors that can occur.

We designed three different visualizations of the precise data to understand whether they affect the bias (RQ2). The baseline condition was to show only the precise data (Figure 2b). We could also show participants both the approximate (Figure 2a) and precise (Figure 2b) visualization

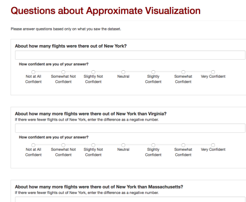
1. http://datahub.io/dataset/open-flights



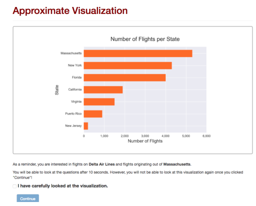
(a) Instructions for the study in general.



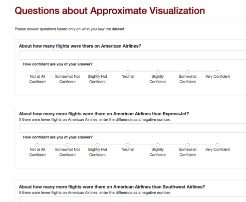
(b) First approximate visualization (here airlines).



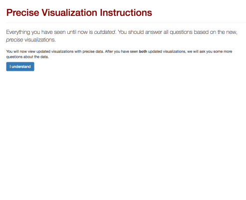
(c) Questions about first approximate visualization.



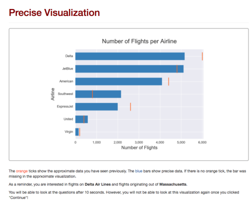
(d) Second approximate visualization (here states).



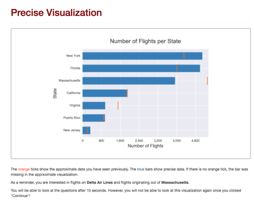
(e) Questions about second approximate visualization.



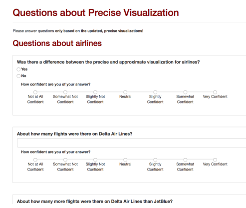
(f) Instructions for the precise visualizations.



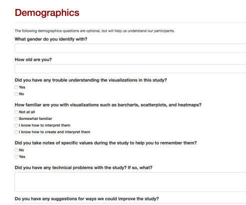
(g) First precise visualization (here airlines and difference visualization).



(h) Second precise visualization (here states and difference visualization).

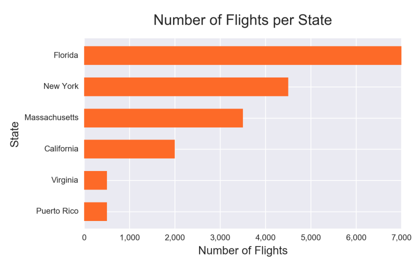


(i) Questions about both precise visualizations.

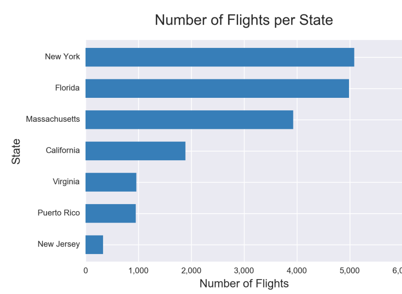


(j) Questions about the demographics.

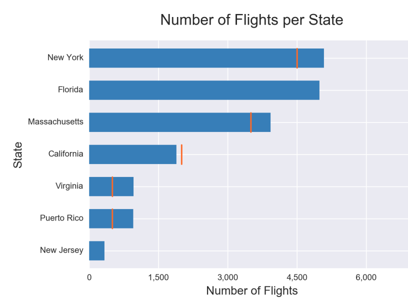
Figure 1. A possible experimental flow with two visualizations about airlines and states with a precise visualization that shows the difference between the approximate and precise data.



(a) Approximate visualization based on 0.2% of the data. Note that “New Jersey” is missing and that “Florida” is the state with the most outgoing flights.



(b) Visualization of precise data.



(c) Precise with difference visualization. The orange ticks show the user the approximate data they have previously seen.

Figure 2. Visualizations of approximate and precise data of the number of flight per state. (c) highlights the difference to the approximate data.

side by side. This method reminds participants of what they have previously seen so that they can clear their memory from approximate results. Lastly, we could overlay the approximate data as small ticks on the bar chart that shows the precise data (Figure 2c).

3.3 Measurements

3.3.1 *Independent Variables*

In our study, each participant was asked to focus on a particular airline and state. All questions were about the one airline and state directly or comparisons with the selected airline or state. The state and airline to focus on were picked at random from the top three airlines and states and remained the same throughout the study.

We evaluated three techniques for presenting the precise visualization (Section 3.2.2). Our control condition presented the precise data as both the approximate and precise data. We further tested presenting the approximate and precise visualizations side-by-side, and a visualization highlighting the difference to approximate data on the precise visualization (Figure 2c). The difference visualization aims to highlight the errors of the approximation as done in Pangloss [2].

We varied how much of the dataset was sampled for each approximate visualization. We tested sample levels of 0.1%, 0.2%, 0.5%, 1%, and 10%, with an additional control condition of 100% (e.g., the approximate visualization was identical to the precise visualization). Smaller samples tended to have greater difference from the precise visualizations. For each approximate visualization, we recorded the difference from the precise visualization relative to the questions we asked. For example, if a participant was asked “How many flights were there on Alaska Airlines?”, we record the difference in number of flights between the approximate and precise visualization.

In this study, we randomized the order that participants saw the two visualizations in; some saw the visualization for airlines first, others saw the visualization of states first.

We used the same order for the approximate and precise visualizations. We also randomized approximate sample, focus state, focus airline, and study condition for each participant. As previously stated, the sample level of the approximate visualizations varied by visualization.

3.3.2 Dependent Variables

We asked three types of questions about each precise visualization, building on Brehmer and Munzner's task taxonomy [15]. Participants were asked to recall (1) how many flights were there on or out of the focus state or airline, (2) the difference in flights between the focus state or airline and another state or airline in the visualization, and (3) which states or airlines were in the visualization. We asked question (2) three times for each precise visualization with three different comparison airlines. This repetition, as well as the repetition between state and airline, gave us repeated measures for each participant.

Each dependent variable has a "correct" answer (e.g., the data presented in the precise visualization). We normalize our outcome (e.g., the answer participants gave) by the by the value of the "correct" answer, e.g.,

$$\frac{correct\_answer - answer\_given}{correct\_answer}$$

We refer to this value as the **measured bias**.

The experimental conditions inform how much variation there was between the precise and the approximate visualizations. For each question we ask the participants, we compute the difference between the "correct" answer (e.g., the data presented in the precise visualization) and the "approximate" answer (e.g., the data presented in the precise visualization). Again, we normalize this value by the "correct" answer, e.g.,

$$\frac{correct\_answer - approximate\_answer}{correct\_answer}$$

We refer to this value as the **expected bias**. We treat the expected bias as an independent variable in our analysis. We hypothesized that the greater the expected bias, the greater the measured bias would be. Meaning, if an approximate visualization differed substantially from the precise visualization, we expected participant's answers to be close to correct (e.g., a measured bias of near 0), but skew in the direction of the approximate visualization. We anticipated that the measured bias would be less than the expected bias, as participants had seen the updated, correct visualization.

For question (3), the answers were calculated via a Jaccard Difference between the airlines or states in the visualization and the airlines or states in the participant's answer.

3.4 Participants

We recruited participants from Amazon Mechanical Turk (AMT), a common tool to run perception studies for visualizations [16]. Similar to other studies of human perception of visualizations, we restricted participation to people on AMT with at least 1,000 tasks completed and an approval rate above 95%. Participation was restricted to the United States to ensure familiarity with the airlines and states in the dataset. Participants were compensated $2.00 for completing the study.

Table 1
The errors for different questions. The results in parentheses are with outliers removed (sigma clipping with $4 \times sigma$).

Question	Data	Mean error	Standard deviation
type 1	approx	0.117 (0.021)	0.535 (0.035)
type 1	precise	0.099 (0.087)	0.167 (0.130)
type 2	approx	0.803 (0.420)	1.649 (0.528)
type 2	precise	2.075 (0.469)	8.500 (0.453)

We recruited 40 participants for the study. One did not complete the study and 5 participants had to be removed because they said they cheated during the study. They said they wrote down the data in the visualizations, despite the instructions asking them them not to. We used the data from the remaining 34 participants for our analysis.

The average participant was 33.7 years old. 26 participants self-identified as male and 8 as female. All participants reported that they had no technical problems with the study procedures. Only 11% of our participants said they are not with all of these visualization types: bar charts, scatter plots, and heatmaps.

4 Results

4.1 Descriptive Statistics

In this study we wanted to understand whether analysts have a bias in how they recall precise data. Before we analyze the bias between what participants remembered from the approximate and precise visualizations, we look at the relative errors that participants made when they answered questions about the visualizations. Table 1 summarized the results. We expect the errors for the precise data to be larger than the approximate one because it is the first visualization they look at. However, that is not true for type 1 questions because of a single outlier where a participant types in a value that was an order of magnitude off. We rerun the statistics with outliers removed (sigma clipping with values outside of $4 \times sigma$ removed). Now the error for the precise visualization is larger (non-paired t-test because missing data does not allow us to run a paired test any more, p=0.0002). For questions of type 2, the errors and variances were much larger and we don't see significantly larger errors in the precise visualizations (paired t-test with mean as value per participant, p=0.032).

Table 2 summarizes the expected and true biases. For question type 1, both the measured and expected biases are relatively small (e.g., about 10% of the correct values). For question type 2, the measured biases averaged over 100% of the correct values, with an even larger standard deviation. This means that participant answers to question type 2 were off by orders of magnitude.

Participants completed the study (excluding instructions) in around 7:30 minutes (with a standard deviation of 3.27 minutes). As expected, most participants spent more time on the questions than the visualizations and overall the most time on the questions about the precise data.

Table 2
The mean and standard deviation of the measured and expected bias.

Question	Bias Type	Mean	Standard deviation
type 1	measured	0.010	0.195
type 1	expected	-0.011	0.215
type 2	measured	-1.542	5.244
type 2	expected	-0.145	0.691

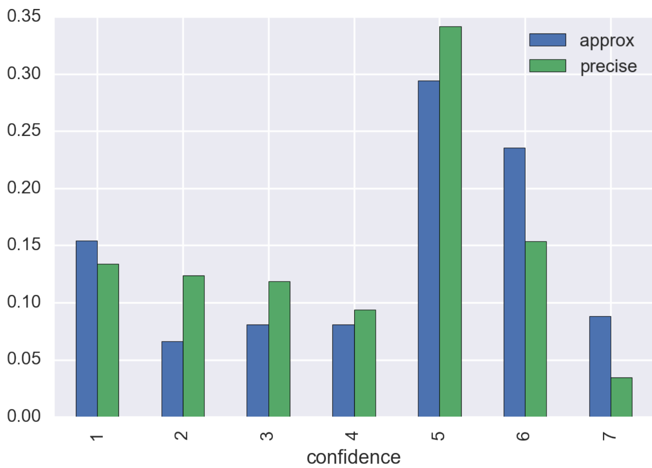


Figure 3. Relative counts of confidence in answers about the approximate and precise data. 1 is the lowest confidence, 7 the highest.

We asked participants to report the confidence in their responses. Figure 3 shows the distribution of relative frequencies grouped by approximate and precise.

4.2 Statistical Analysis

We analyzed our data through linear regression using the `lm` package in R, analyzing the results with a factorial ANOVA (`anova` in R). We hypothesized that multiple factors would impact our outcome metrics, so we used multiple regression in our analyses. We treated each participant as a random effect.

As expected, visualization order, focus state and focus airline had no effect on our results for any of the tests we conducted. For the comparison questions, the comparison state or airline also did not have an effect on our results. None of these independent variables pertain to our research questions, so we discard the terms from our analysis and do not report on them further.

The regression tables of our three types of questions are in Tables 3, 4, and 5. The raw data for our three types of questions are in Figures 4, 5, and 6. We now answer our research questions in turn.

4.2.1 RQ1: Does incorrect approximate data bias responses about correct data?

Our results suggest that viewing incorrect approximate data biases participant responses about correct data for certain types of questions. For all three question types, the expected bias had a significant effect on participant response ($F_{1,61} = 10.767, p = 0.002$, $F_{1,197} = 58.399, p < 0.0001$ for type 2, $F_{1,61} = 78.448, p < 0.0001$ for type 3).

Table 3
Regression table for question type 1, how many flights were there on an airline or state. Baseline is no difference between the approximate and precise visualization, showing only the precise data in the precise visualization.

Study parameter	Est.	SE	t	p
Intercept	-0.033	0.068	-0.482	0.6313
Approx≠Precise	0.214	0.060	0.360	0.7200*
Expected Bias	0.074	0.254	0.290	0.7729
$\text{VisType}_{\text{Diff}}$	0.014	0.055	0.255	0.7999
$\text{VisType}_{\text{Both}}$	0.074	0.058	1.284	0.2041
Expected Bias*$\text{VisType}_{\text{Diff}}$	0.072	0.286	0.252	0.8016
Expected Bias*$\text{VisType}_{\text{Both}}$	0.791	0.317	2.492	0.0154*

Table 4
Regression table for question type 2, comparing how many flights to another airline or state. Baseline is the same as question type 1.

Study parameter	Est.	SE	t	p
Intercept	-2.312	1.066	-2.168	0.0314*
Approx≠Precise	1.745	0.918	1.902	0.0586.
Expected Bias	4.165	0.746	5.584	<0.0001***
$\text{VisType}_{\text{Diff}}$	-0.230	0.850	-0.271	0.7867
$\text{VisType}_{\text{Both}}$	0.576	0.921	0.063	0.9502
Expected Bias*$\text{VisType}_{\text{Diff}}$	-1.474	0.994	-1.482	0.1398
Expected Bias*$\text{VisType}_{\text{Both}}$	-2.247	1.808	1.243	0.2154

Table 5
Regression table for question type 3, which airlines or states were in the dataset. Baseline is the same as question type 1.

Study parameter	Est.	SE	t	p
Intercept	0.217	0.176	1.234	0.2220
Approx≠Precise	0.009	0.033	0.279	0.7815
Expected Bias	0.732	0.202	3.629	0.0006***
$\text{VisType}_{\text{Diff}}$	0.233	0.208	1.119	0.268
$\text{VisType}_{\text{Both}}$	-0.108	0.202	-0.533	0.5961
Expected Bias*$\text{VisType}_{\text{Diff}}$	-0.278	0.236	-1.175	0.2447
Expected Bias*$\text{VisType}_{\text{Both}}$	0.122	0.232	0.526	0.6008

The effect for question type 1 was smaller than the other question types (95% CI -0.435-0.582). This aligns with the finding that the error for this type of question was lower overall. For question types 2 and 3, the sign of the impact of the expected bias was in the same direction of the bias ($95\%CI2.594 - 5.637, 0.329 - 1.135$ for type 2 and 3, respectively). This suggests that people tend to be biased in the same direction as the data they see in the approximate visualization. However, for question type 2, the measured bias was *larger* than the expected bias.

This means that people are reporting answers closer to the approximate visualization than to the precise visualization. One interpretation of this is that there is a significant bias. However, we suspect participants remembered very little about the precise data. Instead we believe this result is heavily influenced by outliers in participant responses.

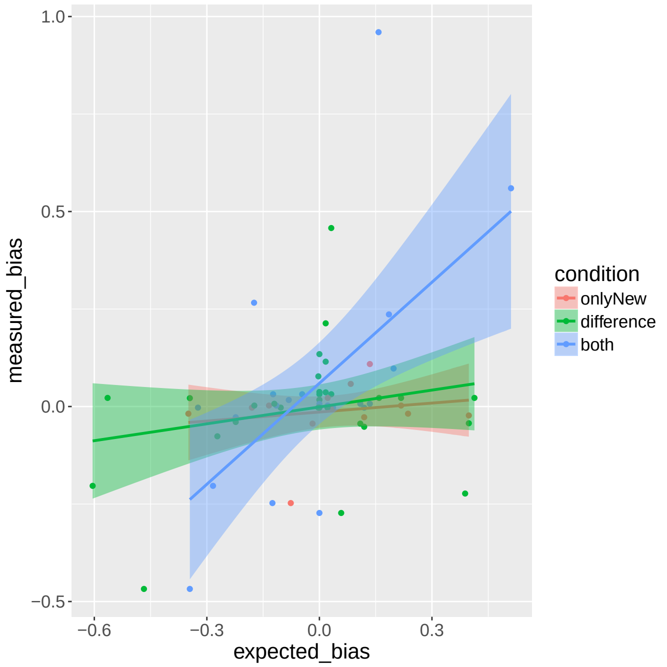


Figure 4. Participant data for question type 1 plotted by measured and expected bias.

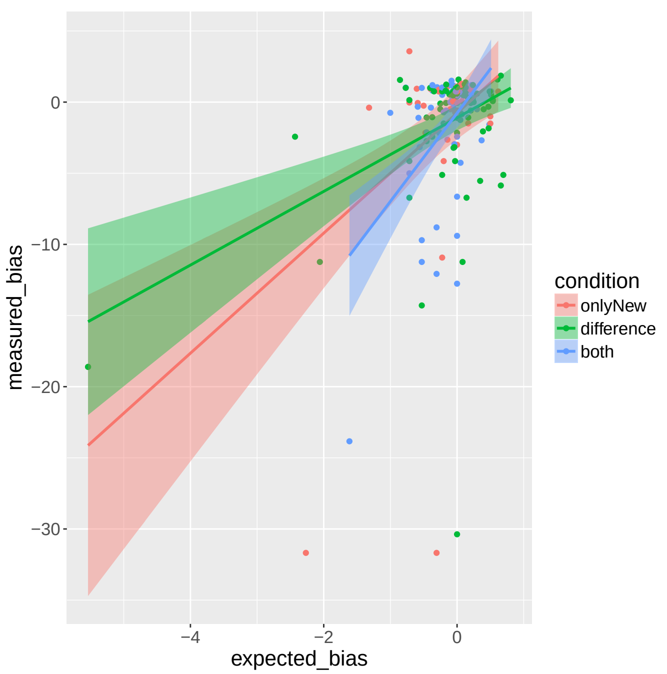


Figure 5. Participant data for question type 2 plotted by measured and expected bias.

4.2.2 RQ2: Does visualization type impact the bias?

Our results fail to reject the hypothesis that visualization type impacts the bias. Condition did not have a significant main effect for any question type ($F_{2,61} = 0.888, p = 0.417$, $F_{2,197} = 0.026, p = 0.975$, and $F_{2,61} = 0.205, p = 0.815$ for types 1, 2, and 3 respectively).

We believe the individual variance between participants exceeded any differences between visualizations. Our results suggest there may be interaction effects between visualization type and the impact of expected bias. However, the variation in participant responses in general made us hesitant to probe deeper into these interaction effects.

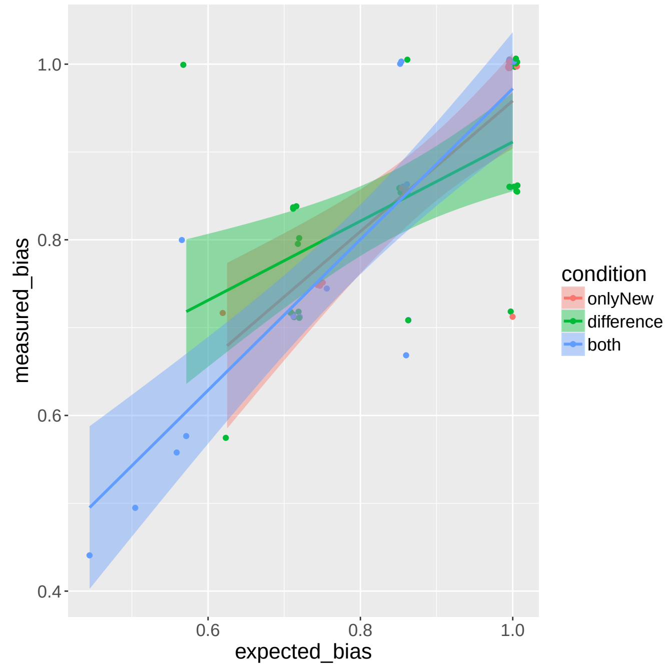


Figure 6. Participant data for question type 3 plotted by measured and expected bias.

5 DISCUSSION

We are a bit unsure why we see the bias we do. On one hand, we predicted we would see a bias in the data. However, we expected the magnitude of the bias would be less than the expected error (e.g., people would be biased only a little by the imprecise visualization). Instead, we seem to be seeing that people are just not remembering the data they saw.

In Section 4.1, we saw that the errors for type 2 questions were much larger than type 1. This makes sense as we asked participants to focus on a particular state and airline and type 1 questions were specifically about that one item. Still, even then errors are very large which could be because participants struggled with the comparison questions. Remember, we asked participants to say how much more flights there were on the focus airline and state and they had to type a negative value if the airline or state we compared to had more flights than the focus airline or state. A more detailed analysis is needed to tease apart why participants made such large errors.

Because errors in the responses from participants were so large, we have little statistical power to tease apart how highlighting differences may reduce biases to answer RQ2. With more statistical power we might also be able to understand the role of how certain participants were in their responses. For each question we asked participants to rate their confidence in their answer. For an analyst it would be fatal to have a large bias and high confidence in their recollection. However, if either the bias is low or the confidence is low, the analysts is likely to not make the wrong decision when it really matters. In our data we did not see significant support for the hypothesis that high confidence correlates with lower biases.

5.1 Threats to Validity

We had no way of ensuring that participants did not take screenshots or notes in order to answer later questions about the precise visualizations. Although we asked participants to not take notes during the study, a few self-reported that they did. We removed these participants from the data but we don't know whether any other participant took notes in secret. Another problematic factor is the large error for type 2 questions (see Table 1). This could be because participants misunderstood the task or because they ignored the instructions.

We used linear regression in all of our statistical analyses. We made this choice because our outcome is numeric rather than categorical. Additionally, we had no reason to anticipate our data would fit a different distribution. However, viewing Figures 4, 5, and 6 shows that although much of the data appears distributed around 0 measured bias, there are many participant responses far away from this value. These responses do not appear to necessarily be outliers, meaning that many people were answering questions similarly wrong.

Linear regression follows a sum-of-squares approach, and is therefore heavily influenced by values far away from the mean. Perhaps a different analysis technique is better suited to our data. We believe our data has systematic problems in how participants are responding to the questions which are unlikely to be corrected by choosing a different statistical test. However, future studies in this domain should ensure that outliers are being treated in a sensible way.

6 Conclusion

The questions we try to answer in this paper in a more general sense are what negative effects seeing a less precise model before a more precise one has on how well people recollect the more precise data. We focused on the specific case where a user only sees one approximate model before the precise one but the findings should generalize to the progressive case because there is nothing inherently different about the precise model. The problem of seeing imprecise models before precise ones is relevant in a number of domains from journalists who may show poll results before the votes, a machine learning expert who builds models, to an economist who presents the forecast of economical development. In this paper, we focused on data exploration in approximate query processing systems with the goal of finding actionable insights that inform how we design such systems.

In this study, we found that exposure to approximate data prior to seeing the precise data biases the observations people make. This has important implications for the design of applications that present approximate results as well as for journalists who report for example polls before votes. In this paper we tried to reduce the bias by making people aware of the approximate result they have previously seen but we did not see a significant reduction of the bias.

In future work, we should focus of usable methods to reduce the bias that can be used by approximate visualization tools. Also, we should also try to understand the relationship between confidence and bias to avoid that analysts make wrong decisions with high confidence. Future work should also investigate how the findings from this paper generalize to other applications.

All material for this study, results, and the analysis are available at https://github.com/domoritz/bias-study.

Acknowledgments

The authors would like to thank Michael D. Ernst for his feedback on the experimental design and writing. They would also like to thank Jessica Hullman and Danyel Fisher for their feedback on the study design and research questions.

References

[1] Z. Liu and J. Heer, "The effects of interactive latency on exploratory visual analysis," *IEEE transactions on visualization and computer graphics*, vol. 20, no. 12, 2014.
[2] D. Moritz and D. Fisher, "Trust, but verify: Optimistic visualizations of approximate queries for exploring big data," in *Submitted to the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2017.
[3] D. Fisher, I. Popov, S. Drucker *et al.*, "Trust me, i'm partially right: incremental visualization lets analysts explore large datasets faster," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2012, pp. 1673–1682.
[4] M. Kay, T. Kola, J. R. Hullman, and S. A. Munson, "When (ish) is my bus? user-centered visualizations of uncertainty in everyday, mobile predictive systems," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2016, pp. 5092–5103.
[5] J. W. Tukey, "Exploratory data analysis," 1977.
[6] S. K. Card, G. G. Robertson, and J. D. Mackinlay, "The information visualizer, an information workspace," in *Proceedings of the SIGCHI Conference on Human factors in computing systems*. ACM, 1991, pp. 181–186.
[7] J. S. Yi, Y.-a. Kang, J. T. Stasko, and J. A. Jacko, "Understanding and characterizing insights: how do people gain insights using information visualization?" in *Proceedings of the 2008 Workshop on BEyond time and errors: novel evaLuation methods for Information Visualization*. ACM, 2008.
[8] J. Nielsen, "Response times: The 3 important limits," *Usability Engineering*, 1993.
[9] S. Agarwal, B. Mozafari, A. Panda, H. Milner, S. Madden, and I. Stoica, "Blinkdb: Queries with bounded errors and bounded response times on very large data," in *Proceedings of the 8th ACM European Conference on Computer Systems*, ser. EuroSys '13. New York, NY, USA: ACM, 2013. [Online]. Available: http://doi.acm.org/10.1145/2465351.2465355
[10] B. Ding, S. Huang, S. Chaudhuri, K. Chakrabarti, and C. Wang, "Sample + seek: Approximating aggregates with distribution precision guarantee," in *Proceedings of the 2016 International Conference on Management of Data*, ser. SIGMOD '16. New York, NY, USA: ACM, 2016. [Online]. Available: http://doi.acm.org/10.1145/2882903.2915249
[11] N. Kamat, P. Jayachandran, K. Tunga, and A. Nandi, "Distributed and interactive cube exploration," in *2014 IEEE 30th International Conference on Data Engineering*. IEEE, 2014.
[12] C. Olston and J. D. Mackinlay, "Visualizing data with bounded uncertainty," in *Information Visualization, 2002. INFOVIS 2002. IEEE Symposium on*. IEEE, 2002.
[13] S. Joslyn and J. LeClerc, "Decisions with uncertainty: the glass half full," *Current Directions in Psychological Science*, vol. 22, no. 4, 2013.
[14] M. A. Borkin, A. A. Vo, Z. Bylinskii, P. Isola, S. Sunkavalli, A. Oliva, and H. Pfister, "What makes a visualization memorable?" *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2306–2315, 2013.

[15] M. Brehmer and T. Munzner, "A multi-level typology of abstract visualization tasks," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2376–2385, 2013.

[16] J. Heer and M. Bostock, "Crowdsourcing graphical perception: using mechanical turk to assess visualization design," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2010, pp. 203–212.