Article

A Comparative Study of Methods for the Visualization of Probability Distributions of Geographical Data

Sanjana Srabanti [1], Carolina Veiga [2], Edcley Silva [3], Marcos Lage [2], Nivan Ferreira [3,*] and Fabio Miranda [1,*]

[1] Department of Computer Science, University of Illinois at Chicago, Chicago, IL 60607, USA; ssraba2@uic.edu
[2] Instituto de Computação, Universidade Federal Fluminense, Niterói 24210-310 , RJ, Brazil; carolinavfs@id.uff.br (C.V.); mlage@ic.uff.br (M.L.)
[3] Centro de Informática, Universidade Federal de Pernambuco, Recife 50740-560 , PE, Brazil; ejs2@cin.ufpe.br
*   Correspondence: nivan@cin.ufpe.br (N.F.); fabiom@uic.edu (F.M.)

Abstract: Probability distributions are omnipresent in data analysis. They are often used to model the natural uncertainty present in real phenomena or to describe the properties of a data set. Designing efficient visual metaphors to convey probability distributions is, however, a difficult problem. This fact is especially true for geographical data, where conveying the spatial context constrains the design space. While many different alternatives have been proposed to solve this problem, they focus on representing data variability. However, they are not designed to support spatial analytical tasks involving probability quantification. The present work aims to adapt recent non-spatial approaches to the geographical context, in order to support probability quantification tasks. We also present a user study that compares the efficiency of these approaches in terms of both accuracy and usability.

1. Introduction

Studying probability distributions is a central task in (visual) data analysis. They are used to model uncertain behaviors present in natural phenomena [1–3], to build predictive models [4], and to describe the behavior/variability present in collections of data, often associated with the processes of sampling and/or aggregation [5,6]. However, effectively visualizing probability distributions presents a big challenge for visualization designers. This fact is due to the infinite variety of possible distributions to represent and the fact that an effective visual metaphor needs to help users reason and perform analytical tasks in probabilistic terms. In fact, as presented in the classical Anscombe's Quartet and in more recent examples by Matejka and Fitzmaurice [7], most common approaches based on statistical summaries, such as mean and variance, can lead to ambiguities and hide important data patterns. Furthermore, when considering distributions in the data analysis process, performing low-level analytical tasks (e.g., comparisons, finding extremes), essential in interpreting a visualization, turns into performing statistical inference by eye [8], which is a complex task even for users with statistical training [9–11]. These problems are even more prominent for thematic geographical data visualizations, which are our focus in this work. The vast majority (if not all) of these visualizations (e.g., choropleth maps) use the position (and possibly the color) visual channel to represent the geographical context, which considerably constrains the design space.

Most of the popular solutions to the problem of visualizing probability distributions of geographical data include using statistical summaries to show central tendency and dispersion measures (e.g., variance). These summaries are used to convey the idea of uncertainty of a value in a particular location and are visualized using, for example, multivariate glyphs, multiple map views, or animation [12]. The focus of these designs was mainly to show the presence/absence of variability/uncertainty; however, they do not address quantification questions such as: *"how likely is it that the values in a particular*

region are larger than a given threshold?" or *"what is the probability of the values in region A being larger than the ones in region B?"* [13,14]. For this reason, users of these visualizations are only able to answer the question *"can I trust this information or not?"* based on the amount of variability, but not to actually *quantify* the likelihood of facts they want to investigate. As stated by several prior works [8,15], quantifying this probability is of great importance in supporting analysts' decision-making process.

To approach the aforementioned problem, in this paper we compare three techniques for the visualization of probability distributions of geographical data, as shown in Figure 1. Concretely, we adapt the frequency-based view from the works of Hullman et al. [16] and Kay et al. [17] to propose two new visual representations called *Hypothetical Outcome Maps* (HOM) and *Distribution Dot Map* (DDM), respectively. Additionally, we adapt the interaction-based approach by Ferreira et al. [8] to propose *Distribution Interaction Map* (DIM). The previous proposals have shown, through user studies, that they can improve users' accuracy and confidence when performing low-level probability quantification of analytical tasks (e.g., cumulative probabilities, direct comparisons between distributions, and ranking) compared to conventional approaches. We highlight that the proposed visual metaphors are variations of choropleth maps. Cognizant of the fact that data sets published by governments and private entities are generally aggregated into geographic units (e.g., neighborhoods, counties, states) and rarely available in higher resolution. While the shortcomings of choropleth maps have been previously pointed out [18], they are still popular in several domains, including public health [19–21].
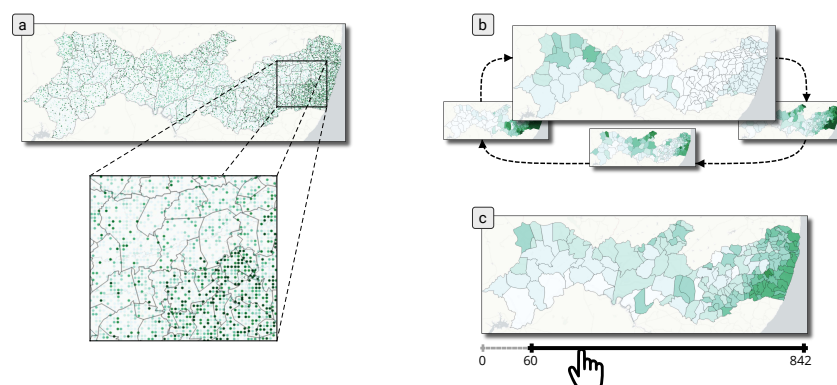

Figure 1. Visual metaphors for the visualization of a probability distribution of geographical data that support probability quantification of analytical tasks tested in this study: (**a**) **distribution dot map (DDM)**, where dots are randomly positioned inside the regions and are colored according to the sampled value, (**b**) **hypothetical outcome maps (HOM)**, where a series of maps are created based on the probability distribution, and then animation is used to cycle through them, and (**c**) **distribution interaction map (DIM)**, where the user can quantify the probability distribution using interactive widgets and annotations.

We also report the results of a user study that compared these techniques. The results show that automating the calculation of the uncertainty quantification in tasks through visual interactions can give the user better precision results in a relative error, time, and confidence (when the visualization supports the interaction for the desired tasks). Even in the presence of factors such as distance between regions on the map, automatically calculating the data quantification does not negatively impact results, which occurs in techniques that do not support interaction.

This paper is organized as follows: Section 2 covers related work in visual analysis of probability distributions, considering both non-geographical and geographical cases; Section 3 introduces the analytical tasks and visual metaphors used in this study; Section 4 presents our user study; Section 5 discusses our results; and Section 6 presents limitations of our study, conclusions and future work.

2. Related Work

In this section, we cover the general problem of representing probability distributions for non-geographical data, as well as approaches for the visualization of geographical distributions.

2.1. Visual Analysis of Probability Distributions: Non-Geographical Case

Designing techniques for visualizing probability distributions have long been an important research problem for general abstract data [22]. One of the main applications of these techniques is to represent uncertainty. Depicting uncertainty is essential for decision-making [17] and at the core of recently proposed techniques to support interactive visualization of large data [4–6]. The general solutions to visualize probability distributions (often sample distributions) are static representations such as error bars, confidence intervals and box plots. These are based on statistical summaries such as moment statistics like mean and variance (e.g., error bars and variants) [11], spread measures such as entropy [23] and order statistics (e.g., box plots) [24]. Such approaches present major drawbacks. Not only can they be ambiguous, in the sense that many distributions have the same visual representation [6], but they also are often hard to interpret, even for trained individuals [25]. As a result, these techniques are designed to show the *presence* of variability in the distribution (i.e., uncertainty), but do not directly support the quantification of uncertainty for analytical tasks [8]. This is not ideal, since the major goal of the visualization should not be just given intuition of how certain a particular estimate is (which is just one of the aspects of a distribution), but rather supporting analysis and decision-making in these scenarios. Blumenschein et al. [26] proposed a technique called v-plots, which builds a hybrid graph consisting of different layers corresponding to different traditional plots. The rationale behind this technique is that real-world analysis scenarios involve multiple analytical tasks, and therefore require multiple plots to support them (which is similar to the premise used by Ferreira et al. [8]). While this approach can mitigate some of the limitations in particular techniques, the approach is difficult to generalize to the geographical case. As we comment on later, building a hybrid map approach to better support different tasks is an interesting direction of research.

Recently proposed techniques try to solve these issues by relying on the concept of *implicit uncertainty* [27] in their visualizations. In fact, instead of explicitly encoding statistical summaries or aggregations, these visualizations use the human perception to derive representations that depict individual samples of these distributions. Therefore, the interpretation of uncertainty in this case is based on a frequency-based view of probability. As discussed in the work of Nguyen et al. [28], and references therein, a large body of research in psychology points to the efficiency of this approach in terms of cognitive effort, accuracy, and time to respond to analytical tasks. For instance, Hullman et al. [16] proposed a technique called *Hypothetical Outcome Plots* (HOPS). This consists of showing multiple samples of the distributions being analyzed through animation. The authors validated this proposal through a user study via Amazon Mechanical Turk. They concluded that for quantification tasks like estimating mean or comparing/ordering distributions, HOPS is able to significantly improve users' accuracy compared to traditional error bars and violin plots. A particularly positive aspect of this technique is that it is relatively easy to generalize to different visualization scenarios, particularly for the geographical one. However, a significant limitation is that the tests were only done with a limited number (up to three) of distributions. Another example is the work by Kay et al. [17], which proposes the *Quantile Dot Plot*. This technique is a modification of the classical dot plots, which can be viewed as a discrete analogous of histograms because stacked individual observations can be seen instead of bars. This technique was evaluated in the paper mentioned above and in the work of Fernandes et al. [15]. Quantile dot plots were compared against more common visualizations like strip plots, cumulative density function (CDF) plots, and probability density functions through a Mechanical Turk user study emulating a scenario of predicting a bus' arrival time. They found that quantile dot plots perform nearly as well as CDF plots, but outperform other plots. Furthermore, through an incentivized experiment, the

paper shows interesting evidence that the ability to better quantify probabilities from the visualization leads to better decision-making. The discrete nature of these frequency-based visual designs makes probability quantification intuitive and straightforward even for people with no training in statistics. Basically, they estimate the number of times (frequency) a certain event happens (e.g., frames in the animation or dots in the quantile dot plot).

Ferreira et al. [8] proposed a different approach to quantifying probability in analysis tasks, focusing on the scenario of progressively estimated distributions. In this case, distributions are estimated from samples that evolve over time to provide interactive visualization for large data sets. For this reason, following a set of proposed guidelines, they suggested using interactive annotations alongside standard visualizations. An annotation is created to support each analytical task considered (cumulative distribution estimation, distribution comparison, ranking, and quantifying the probability of extreme values), and each of them enables uncertainty quantification without the use of inference by the eye. This was considered a good compromise since in progressive analysis scenarios, users do not have enough time to judge a particular state of the distributions, and also animation is usually already used for the evolving estimates. This approach was evaluated through an in-person user experiment to compare it against bar charts with error bars. The results showed that it improved user confidence while not causing an increase in time of judgment. The negative aspect of this approach is the need to design visual strategies for each task.

The present work aims to generalize and compare both frequency-based and interactive approaches to the geographical visualization scenario.

2.2. Visual Analysis of Probability Distributions: Geographical Case

Summarizing data distributions is at the core of thematic cartography. Widely used techniques such as choropleth maps and cartograms use central tendency statistical summaries, commonly mean or median, to depict spatial patterns in geographically tagged data. In the vast majority of these cases, however, many aspects of distributions are ignored [29], which, similar to the non-geographical case, can harm the analysis process.

Visualizing geographical distributions is often considered in uncertainty visualization since uncertainty in the data is often modeled through probability distributions. Previously proposed solutions are discussed in great detail in the surveys by MacEachren et al. [12] and Ślusarski and Jurkiewicz [13]. According to these surveys, proposed solutions vary in the data type considered, uncertainty type (positional, temporal, or attribute value), and visualization strategy (static, dynamic, and interactive). In the present work, we will consider the different strategies concerned with attribute value uncertainty of geo-tagged numerical data since this is the most popular use case for thematic mapping and geographical data exploration.

Traditional solutions employ a static strategy for geographical data, analogous to error bars in the non-geographical case. In other words, they visualize the central tendency of a distribution and its variance. To do this, proposals use visual channels such as transparency [30], uncertainty contours [31], texture [32], noise annotations [33], bi-variate color scales [34,35], or multiple views [12] (one for central tendency and other for variability). Another approach is to employ glyphs [12,36,37], with the advantage of having more design freedom to depict the distribution. However, the placement for glyphs is problematic, often resulting in overplotting. Again noticing the similarity of the non-geographical case, we observe that these solutions are not designed to support the probability quantification of analytical tasks, but rather often only show the presence/absence of uncertainty. One interesting exception among static methods is Dot Maps. These traditional techniques consist of placing symbols on a map to represent the spatial distribution of discrete geographic data [38]. A popular application of this map type is the Racial Dot Maps [39], which show the ethnic distribution in discrete geographical regions. The dots are colored according to the corresponding category (race) and are randomly placed inside each region in a way that the proportion of dots of a particular color within a region corresponds to

the proportion of the race in that region. We emphasize that Racial Dot Maps represent categorical proportions, not uncertainty per se.

There are many possible variants of dot maps in terms of dot placement and size. We refer to Chua and Moere [40] and Hey [41] for discussions on these variants. We believe that the discrete nature of these maps brings them close to the frequency-based approach of dot maps and, therefore, could be used to visualize continuous distributions (see Section 3.2).

Dynamic approaches use animations to represent uncertain information. For example, Evans et al. [42] proposed the use of a "flickering" strategy in which an animated map alternates between showing the data (central tendency) and only showing areas with low uncertainty. A similar approach was used by Kardos et al. [43]; however, the flickering was defined to be between the data layer (central tendency) and the uncertainty layer, which showed the standard error of the distribution in each location. While these techniques were found to have comparable performance to static techniques (in terms of user accuracy and viewing time) and user-rated helpfulness, the animation was found to be bothersome for some users. We again notice that, as in the static displays, these approaches make quantifying the probability of analytical tasks a difficult problem. Ehlschlaeger et al. [44] proposed to animate realizations of a model of land elevation. They proposed a method for ordering the different frames (model realizations) and producing intermediate frames to have a smooth animation. Although their method can be seen as a geospatial HOP (and therefore similar to our HOMs approach), the ordering of the frames requires users to watch the full animation to have an unbiased view of the data. To remove this limitation, HOMs use independent random samples similar to HOPs. More recently, Cybulski [45] presented a study of spatial pattern and temporal trend recognition on animated choropleth, highlighting the need of specific visual strategies.

Early interactive methods focused on allowing the user to control, via interactively defined thresholds, which portions of the data should be hidden or represented less clearly [46,47]. Furthermore, Cliburn et al. [48] visualize uncertainty in an ensemble of climate models. For this, they use both glyphs and color mapping strategies. Recognizing that the value of uncertainty may be hard to read, their system allowed users to perform selections (point or region) and see the associated uncertainty values. This can be thought of as an intermediate approach between static and animation methods described above. More recent interactive methods tie the interaction design with the probability quantification of analytical tasks performed during the data exploration. For example, the system proposed by Alberti [49] supports the quantification of the probability of the values of spatial distributions inside user-specified intervals. To do this, the user specifies the interval with a slider, and quantified probabilities are presented as circular glyphs, which are placed at the center of the cells of a spatial grid. In a different example, Pahins et al. [6] proposed an interface to support the task of quantification of quantiles and cumulative distribution function (CDF). Using a slider, the user specifies a query parameter (either a value between 0 and 1 for quantiles or a value to evaluate the CDF) using a slider. The visualizations show the quantiles or CDF values corresponding to each region in a choropleth or a heatmap. We highlight that this approach and its consequences have been relatively unexplored in the visualization of geographical data. In fact, to the best of our knowledge, in the geographical domain, no work followed the same line as in the non-geographical case (for example [8,11,16]) to assess the efficiency in terms of low-level analytical tasks focusing on probability quantification. Rather, previous evaluations of geographical uncertainty visualizations are often concerned with quantitative tasks of finding classes (levels) of uncertainty, distinguishing between these levels, and ranking them [13,14], such as "*which region is more/less uncertain?*".

3. Background And Preliminaries

Analyzing data through its probability distribution is an increasingly common approach in several domains, either because the underlying captured phenomenon is uncertain and modeled as a probability distribution or because operations (such as aggregations)

applied to the data result in a distribution. The latter is the scenario we will use here as a running example. We start our discussion with a brief description of the mathematical tools used in modeling this scenario.

A random variable (RV) represents a quantity whose value depends on the outcome of a random phenomenon. We will represent RVs using upper case letters such as X, Y, and Z and constant values that are not random with lower case letters such as a, b, and c. A RV has an underlying distribution that encodes the probability of the associated RV assuming a given value (discrete case) or being inside a given range of values (continuous case). We formally define an aggregation as a derived value (or statistic) that summarizes one aspect of the distribution of a RV. Aggregation is an essential tool to build visual summaries of data and, therefore, for interactive visual analytics systems [50–54].

In this paper, we are concerned with the analysis of distributions of numerical geographical attributes. We focus on the scenario in which we have a discrete set of spatial regions and, associated with each one of them, we have a RV modeling a collection of data points belonging to the region. For example, Figure 2 presents the mean number of monthly taxi trips in Manhattan for a period of one year. While the aggregated map only shows a statistical summary (e.g., mean, sum) (a), there is an underlying monthly distribution of the number of trips over the entire year (b).
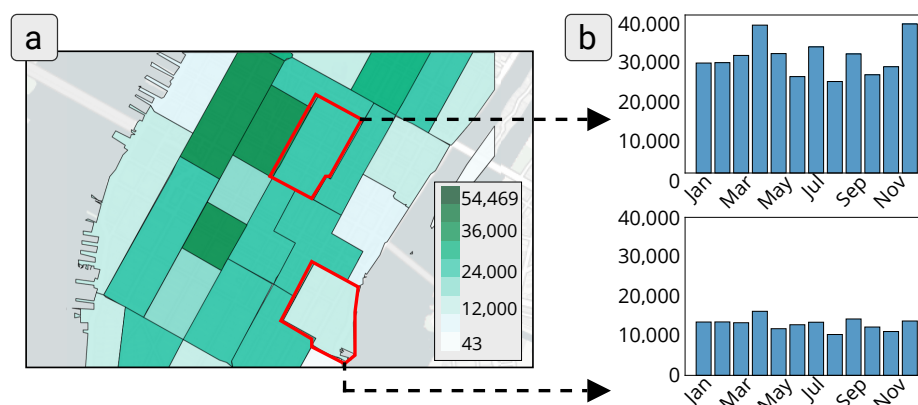

Figure 2. (**a**) A choropleth map showing the number of taxi trips in each neighborhood of Midtown Manhattan (NYC). (**b**) Each of these neighborhoods has an associated monthly distribution of the number of taxi trips, as highlighted for the regions of Midtown Center and Kips Bay (**b**).

3.1. Analytical Tasks

While visual summaries such as Figure 2a use aggregation to reduce the amount of data that the user has to analyze, it can hide important patterns as well as give the user a false impression of certainty/precision of the presented value. Therefore, when facing such cases, it is important for the analyst to visualize more details of the underlying distributions and perform analytical tasks that extract information from these distributions. We draw upon past literature [8,16,55] to define a set of analytical tasks that will guide the design of the visual metaphors, the formulation of hypotheses about their performance, as well as the design of a user study to compare visualization techniques.

Amar et al. [56] present a series of analytical tasks that are common in the visual analysis process. Examples of such tasks are: retrieve value, filter, and find the minimum and maximum. Andrienko and Andrienko [55] studied these tasks and their use for the exploration of geographical data, which we use as the main reference for task selection. Their work, however, did not take into account the use of probability distributions, resulting in a more straightforward translation between task and visual representation. Ferreira et al. [8] considered the problem of visualizing probability distributions for non-geographical data and translated a subset of the tasks proposed by Amar et al. to a probabilistic setting. This approach consisted in turning the deterministic tasks on inferences that estimate the probability of occurrence of certain facts. For example, a *retrieve value* task is translated

to the probability of a sample from a given distribution to be smaller than a predefined value (i.e., $P(X \leq c)$). They also discuss a version for the task that involves estimating the probability of a sample to fall within a given interval (i.e., $P(c_1 \leq X \leq c_2)$). Similarly, a comparison task, which in a purely deterministic setting is simply comparing two values, is translated to the task of evaluating the probability of a sample in distribution X_1 to be smaller than a sample in distribution X_2 (i.e., $P(X_1 \leq X_2)$). A similar approach was used by Hullman et al. [16] to evaluate visualization of distributions, again for non-geographical data. Given that, we can define the following analytical tasks for the probability distribution of geographical data:

Retrieve value. Given a geographical region R, estimate the probability of a sample to be in a given interval. The interval can be either bounded $P(c_1 \leq X_R \leq c_2)$ or unbounded, in which case we have either $P(X_R \leq c)$ or $P(X_R \geq c)$)), where X_R is the RV describing the distribution of the data associated with R. Considering the example in Figure 2, this task estimates the probability of having the number of taxi trips in a given region greater than a certain threshold.

Find extremes. For all geographical regions R_i, find the one that minimizes (or maximizes) the probability of a sample to be within a given interval. More clearly, given an interval I (of one of the types described above), find the region that minimizes $P(X_{R_i} \in I)$). In Figure 2, this task finds the city region that has the highest probability of having a sample, for example, greater (or lower) than a certain threshold.

Compare distribution. Given a pair of geographical regions, R_1 and R_2, estimate the probability of a sample in the first region being smaller than one of the second region, i.e, $P(X_{R_1} \leq X_{R_2})$). In Figure 2, this task quantifies the probability of having, for example, fewer taxi trips in a given city region compared against another one.

Estimate mean. Given a region R, estimate the mean (i.e., expected value) of the associated distribution of associated RV, X_R.

The previous tasks were selected because they are common in the analysis of spatial data with uncertainty. It is important to notice that while some of the tasks are in some sense local (i.e., depend only on one or two regions of the map), the spatial disposition of the regions on the map will impact how efficiently they can be performed. Put in a different way, these tasks could, in principle, be interpreted as non-spatial, but in reality, they are very important for map interpretation [55] and to support high-level tasks. In addition, while the estimate mean task could be interpreted as a straightforward task (since one could directly visualize the mean on the map), we believe that an efficient display of distribution should allow the user to estimate common summaries that could be useful during data exploration. Visualizations that try to directly convey multiple summaries can become crowded and hard to interpret. For this reason, similar to the study performed by Hullman et al. [16], we decided to use the estimate mean task in our evaluation.

3.2. Visual Metaphors

As previously discussed, the most direct approaches to visualize geographical data, such as choropleth maps and their variants (that represent measures of dispersion), do not directly support the tasks described in Section 3.1. The consensus in different domains is that visualizations need to either depict the entire data (which is generally infeasible) or *give more direct access to the distributions* than just mean and variance [57–59]. To achieve this in the case of geographical data, we propose to adapt recent ideas used in the context of non-geographical data. Next, we highlight each one of these adaptations, focusing on how they can be applied to the visualization of geographical data. The visualizations are available at https://vis-probability.github.io/demo/ (accessed on 12 July 2022).

3.2.1. Distribution Dot Map

The first technique is an adaptation of the Racial Dot Maps (RDM) to depict numerical distributions, which we call *distribution dot map* (DDM). Figure 1a shows an example of a *DDM*. Similar to RDMs, in order to visually represent the probability distribution P_i of a

geographical region A_i, this technique draws dots randomly positioned in a grid inside the polygon that defines the region. There are two important observations here. First, we chose to use randomly positioned colored dots to represent the quantitative values of the distribution, since, (1) this is a common approach used in RDMs and different from ordinary (non-spatial) dot plots, and (2) there is no standard way to organize the dots spatially. We believe that the choice of an arbitrary ordering could cause the appearance of undesired geographical features. Second, the grid-based approach used here, similar to the dot placement described by Chua and Moere [40], avoids overlapping dots that are common in ordinary RDMs. As discussed by Chua and Moere, density estimation is negatively impacted by overlapping dots. Lucchesi et al. [60] used a similar placement approach to create a region pixelization strategy to display uncertainty.

Each dot in a DDM corresponds to a distribution sample and is colored according to the sampled value. The larger the probability of a given value in the distribution, the more dots will have the color associated with that sample value. We chose the use of color instead of dot size to represent the distribution values to, again, avoid overlapping dots as well as to make dot density estimation easier. In addition, in a DDM, the number of dots within a region is proportional to the area of the geographical region. We notice that, while having the same number of dots in all the regions would be desirable to more easily support comparisons, the diversity of region sizes in geographical visualization can prevent the representation of complex distributions in larger regions due to smaller regions present on the map.

We see DDMs as an extension of the frequency-based approach provided by *dot plots* [17] to the spatial context. For example, notice that this visualization turns the process of performing the analytical tasks previously described into counting/estimating the number of dots inside one (or more) region(s) with a certain color. This task is similar to the dot density estimation task performed in RDMs.

3.2.2. Hypothetical Outcome Maps

The second technique that we propose considering the spatial context is *hypothetical outcome maps* (HOM), based on the *hypothetical outcome plots* [16]. This consists of an animation showing a series of choropleth maps displaying samples (hypothetical outcomes) of the distributions associated with the represented regions. The sampling can be achieved in different ways; without apriori correlation/dependency, samples can be drawn independently from the different distributions P_i. However, when there is an apriori correlation, it is possible to sample the entire map as an outcome of a joint distribution of all P_i. For instance, in the example of taxi trips in NYC, there is a link between the number of trips and the months of the year. The resulting animation shows random samples of the joint distribution over the twelve months of the year. One important point is that the color map function is fixed throughout the animation, in order to make comparison across frames possible. Hullman et al. [16] suggested the use of 400 ms as the transition time between frames of the animation; however, in a pilot study, we found that users had some difficulties with fast transitions. We, therefore, set the transition time to 660 ms as a good compromise for analysis. Figure 1b shows an example of an *HOM*.

3.2.3. Distribution Interaction Map

The last technique is based on the proposal by Ferreira et al. [51]. This technique consists of fixing the (preferably small) set of tasks that will be supported and, for each one of them, designs interactive widgets/annotations that support automatic quantification of the probabilities in the selected tasks. We call this technique *Distribution Interaction Map* (DIM). To support the retrieve value and find extreme tasks, the interface shows a slider where the user can define an interval $[c_1, c_2]$. The system then uses this interval to quantify the value of $P(c_1 \leq X \leq c_2)$ via a sample drawn from the distributions. The resulting probabilities are visualized as a choropleth map, as shown in Figure 1c. We notice that this

approach has been used in practical visual analytics systems [6,49]; however, they did not evaluate this technique nor explicitly support other tasks.

For the comparison task, we designed an interaction similar to the one proposed by Ferreira et al. [8]. In this case, the user double clicks on a region R, and all the other regions in the map will be colored according to $P(X_A \leq X_i)$, where X_A is the RV associated with the region R and X_i is the RV associated with each of the other regions.

To test how well this technique supports the estimate mean task, we decided not to create a specific interaction, unlike for the previous task. In theory, using the previous interaction, it is possible to estimate the mean with $\int_{[0,+\infty)} P(X > t) dt$. However, we expect that users will report difficulties in doing so.

4. Evaluation

We conducted an in-person user study to evaluate the effectiveness of the three techniques described in Section 3.2. The methodology of our experiment is similar to the ones used by Ferreira et al. [8] and Hullman et al. [16]. In our study, each participant was randomly assigned to one of the three visualization conditions (i.e., each user was only exposed to one of the techniques). The user study was composed of four parts. In the first part, participants were asked to complete a questionnaire with personal information such as age, gender, and educational background. In the second part, we presented the participant with a short tutorial describing the technique they would use in the remainder of the study. At the end of the tutorial, the participants were asked to answer a simple question comparing two distributions in two different regions, in order to evaluate if they understood the basic functionalities of the visualization technique. After successfully answering the tutorial question, participants answered a set of questions on the tasks described in Section 3.1 facing stimuli that simulated data analysis scenarios, using both synthetic and real-world geographical data sets. For each question, we recorded quantitative information to evaluate each of the visualization techniques.

Experiment apparatus and data collection. The participants performed the experiment using a web-based application containing all the questionnaires and quantitative tasks. We implemented the three visualization techniques using JavaScript and D3, and Leaflet and Turf for mapping and spatial analysis. The three interfaces display a colorscale on the bottom right side of the map. We performed a pilot study before the user study to adjust interface parameters, including HOM animation speed, and number of dots in DDM.

4.1. Study Design

The study consisted of a main questionnaire in which the participant answered a series of 26 questions based on two real-world spatial data sets (13 questions per data set), covering the four analytical tasks presented earlier in the paper (for each data set: 3 questions to retrieve the value, 2 questions to find extremes, 7 questions to compare regions, and 1 question to estimate the mean). The number of questions per task was different due to possible variations of the task and/or independent conditions that we wanted to test. For example, in the retrieve value task, we included variants that asked the user to estimate, for a given region, the probability that samples are in a bounded interval or, also, either above or below a threshold, totaling to 3 possible variants. For the task of comparing distributions, we selected 3 factors for which we wanted to measure their influence on the task, namely, region size, region distance, and distribution variance. We have 2 questions associated with 2 possible values of the factor and 1 extra question that did not consider these factors, totaling to 7. Some of the scenarios that tested the 3 factors mentioned above are shown in Figure 3: (a) the size of the regions, (b) the distance between regions, and (c) the distribution variance in the regions. All of the examples of questions were manually chosen by the authors to present reasonably challenging problems. In particular, for each comparison question, we selected the regions based on these variables (i.e., small and large regions, nearby and faraway regions, low and high variance regions).
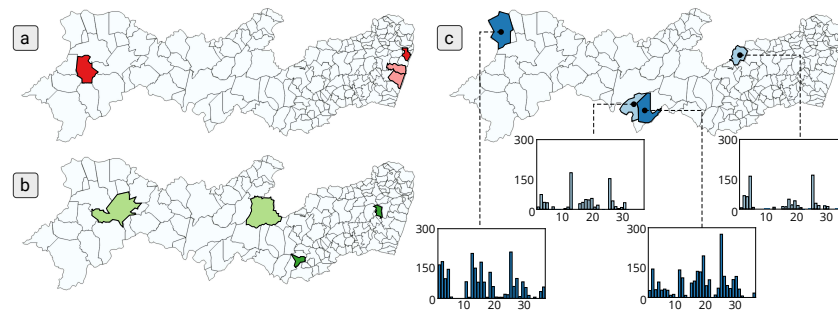
Figure 3. Variables considered in the comparison distribution questions of the user study: (**a**) distance between regions, (**b**) region size, (**c**) data variance. Examples were manually selected to compare these factors. In each scenario, we paired regions with the same shade of color (dark–dark, light–light).

For each question, participants were shown a map with one or more highlighted regions (as in Figure 2a) and were asked to perform estimations based on these regions. For each question in the tasks (presented in random order), we recorded the user's answer, time used to respond, and also the self-reported confidence in their answer. The answer was a numerical value that represented the estimated probability (from 0 to 100) typed by the participant. The only exception was find extreme tasks, for which the participants chose a region name from a drop-down menu containing all the regions sorted in alphabetical order. The confidence was measured on a five-point Likert scale from "completely uncertain" to "completely certain". Table 1 shows some examples of questions for the tasks in the user study.

Table 1. Example of questions in the user study.

Analytical Task	Question
Retrieve Value	What is the probability to rain less than 100 mm in A?
Find extremes	Select the region with the highest probability to have a number of taxi trips greater than 34,000.
Compare distributions	What is the probability to rain more in A than in B?
Estimate mean	Estimate the average number of taxi trips in A.

Data sets. To generate the actual test scenarios, we used two different real-world data sets. The first data set is composed of rainfall indices collected in the 184 cities located in the Brazilian state of Pernambuco, with measurements for each month between 2016 and 2018 (36 values per city). The second data set is comprised of taxi trips in 29 regions of Midtown Manhattan, a borough of New York City (NYC). The data is composed of 7,991,059 taxi trips and was aggregated by month, resulting in a distribution of 12 values per region. These data sets were chosen to simulate real data analysis scenarios involving geographical distributions.

Participants. We recruited 78 people to participate in the user study, 52 males and 26 females. They were randomly assigned to three groups: DDM, HOM, and DIM, so that each group had 26 participants. All participants had at least a bachelor's degree, and 46 of them had a graduate degree. Furthermore, all participants attended at least an undergraduate course in statistics. From the total number of participants, 49 of them reported having experience in statistics (12 in the DDM group, 20 in the HOM group, 17 in the DIM group), and 54 of them reported having experience in visualization (16, 17, and 21 for each group, respectively). The participants became familiar with the study regions in the tutorial section. None of them were color blind or had previously used the techniques.

4.2. Hypotheses

Considering the features of tasks discussed in the previous section and the visual metaphors used, we formulated a set of hypotheses that we want to validate using the collected data. We discuss these hypotheses in the following.

Hypothesis 1. *DIM is the technique with higher accuracy for the tasks of retrieve value, find extremes and comparison.*

We can consider that answering each of the tasks involves two steps. The first one (called the estimation step) extracts some numerical properties from the visualization, for example, a probability quantification involving one or more distributions, or estimating the mean of a given distribution represented in the visualization. The second step (called the consolidation step) uses these properties to form the answer to the questions asked.

While all three techniques, by design, try to make probability quantification more intuitive and easy, they impact these two steps differently. For the tasks of *retrieve value*, *comparison*, and *find extremes*, the interaction strategy makes the extraction step straightforward since the estimated probabilities are directly represented using ordinary choropleth maps. We noticed that the cognitive load involved in this step is much higher since the user will have to rely on memory (in the case of HOM) or to estimate the proportion of a certain collection of dots inside a particular region (in the case of DDM). Therefore, we expect that the DIM technique will better support the tasks above compared to the other techniques. While it may seem strange/unfair to compare it against the other techniques, in this case, we believe that this is important for two reasons. First, it serves as a baseline for us to evaluate how well both the HOM and DDM would perform compared to the *ideal case*. Second, testing this hypothesis will help us evaluate the usability of this technique. Finally, we expect that HOM and DDM will have similar performances in terms of accuracy for these tasks. We think that the cognitive effort involved in evaluating the animated scenario will be matched by the one involved in estimating the density of points in a region with a certain value.

Hypothesis 2. *DIM is the technique with equal or smaller completion time for the tasks of retrieve value, find extremes, and comparison.*

Considering the discussion above, we expect that the DIM technique will outperform the other techniques in terms of time spent on answering the questions for the same set of tasks mentioned above. Again, the goal here is to use the DIM approach as a baseline for the effectiveness of the other methods and test its interaction usability.

Hypothesis 3. *DDM and HOM have lower confidence than DIM for for the tasks of retrieve value, find extremes and comparison.*

We expect that the cognitive load involved in interpreting the animations (HOM) or counting the possibly large number of dots (DDM) and remembering the associated values will be tiring and error-prone. For these reasons, we expect that the users will feel more confident in reading the values directly from the quantified probabilities given in the DIM approach.

Hypothesis 4. *HOM will better support (in terms of accuracy and time) the estimation of the mean.*

The estimation step is not well supported in the DIM approach. As discussed in Section 3.2, estimating the mean in this approach is challenging. Furthermore, we expect the estimation step in the DDM approach to be error-prone, given the need to accumulate

sums and count the dots. The randomness present in the HOM animations should be a powerful tool for estimation in this case.

Hypothesis 5. *Distance between polygons, size of the regions, and magnitude of distribution variance of each region negatively impact accuracy in the comparison task when considering the DDM and HOM techniques.*

One of the main aspects of geographical analysis compared to the non-geographical scenario is the spatial dispersion of the visual representations of the distributions. In our scenario, for the tasks that involve the consideration of multiple distributions, say the comparison task, we expect that the distance between the regions will negatively impact both the accuracy and time of the responses for the HOM and DDM techniques.

4.3. Quantitative Results

For the quantitative analysis of the collected result, we first used the Kolmogorov–Smirnov (KS) test to reject the normality of the distribution of the samples in the data for the conditions considered (none of the scenarios tested generated normal distributions according to the test). We then use the non-parametric Kruskal–Wallis (KW) to test if the distributions of *all* tested groups are equal. If there is a significant difference between groups, we use the Mann–Whitney–Wilcoxon test (MWW) to check if the median ranks of *each* pair of groups are equal. For all tests, a significance level of 0.01 was used after using Bonferroni correction to control for Type I errors, i.e., we regard results significant if the associated *p*-value is at most 0.01. To give a more concrete meaning to the accuracy study and make it independent of the data set used, the accuracy is reported as the relative estimation error.

Retrieve Value Task. The participants in the DIM technique registered the smallest relative errors, with a significant difference when compared with the DDM and HOM techniques (Figure 4 (left)).The results provide sufficient support for Hypothesis 1 (H1) in this task. The completion time of this task was, on the whole, shorter between the participants in the DIM group. Additionally, both DDM and HOM have significantly higher outliers than the DIM group (Figure 4 (middle)). The results support H2. The DIM technique, since it allows us to precisely quantify the probability, had a significantly higher confidence score (Figure 4 (right)). More than two-thirds of the participants in the DIM group said they were very confident (5) in their answers, as opposed to less than one-third in the other two groups. This result supports H3.
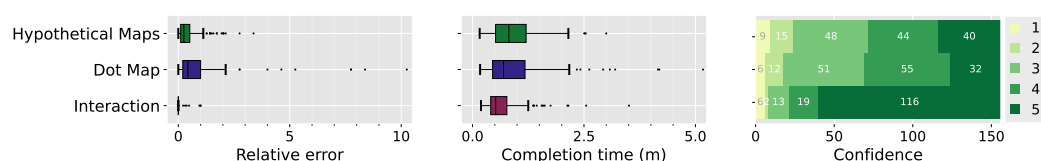

Figure 4. Results of the retrieve value task.

Find Extremes Task. In the find extremes tasks, users have to report the name of the region identified as having the maximum (or minimum) estimated probability, i.e., we want to find the region X, which maximizes (or minimizes) $t_X = Pr(X < c)$, for a given constant c. To numerically quantify the errors in this task, we compute the relative error of the probabilities associated with the region chosen by the participant and the one that is the extreme. In other words, if X is the chosen region and R is the extreme region, the relative error is computed as $|t_X - t_R|/t_R$.

The participants in the DIM group again registered the lowest number of errors, a significant difference when compared to HOM and DDM (Figure 5 (left)). Given that, the results provide sufficient support for H1 in this task. Noteworthy is the difference between the relative error in the two data sets. In the rain data set, participants in the DDM and HOM groups registered similar relative errors. We have a different scenario in the taxi data

set, where the HOM group had a smaller relative error than the DDM group. This can be explained by the fact that the geographical regions in Manhattan are mostly rectangular, with a smaller total area compared to the rain data set, resulting in less space to insert the dots that represent the distribution. Additionally, the taxi data set distribution has minimal variation, which can easily confuse participants with the approximation made in representing the dots on the map.
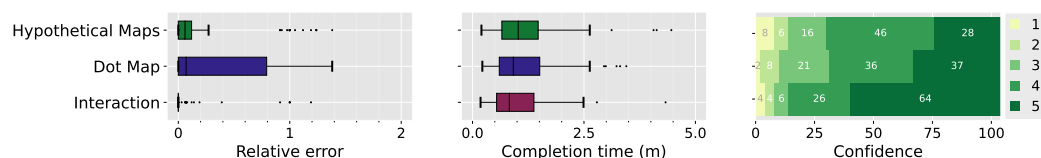

Figure 5. Results of the find extremes task.

The DIM group were significantly more confident in their answers when compared to the other two groups (Figure 5 (right)), supporting H3.

Compare Distributions. The relative error was once again significantly lower among the participants in the DIM group (Figure 6 (left)). This result also supports H1. In terms of completion time, the DIM technique once again showed the lowest values (Figure 6 (middle)), with a significant difference to the other techniques. This is expected, as the DIM technique provides the quantified probability in an automated way while the other two demand a complex mental estimation. The pairwise differences between the three techniques are also significant, for all pairs except the DDM and HOM. This result supports H2. The confidence results for this task showed that participants were more secure using the DIM technique, with over two-thirds using a rating of 4 or 5 (Figure 6 (right)). The KW test showed a significant difference between the three groups, and the MWW test showed a significant difference between pairs of groups. This result supports H3.
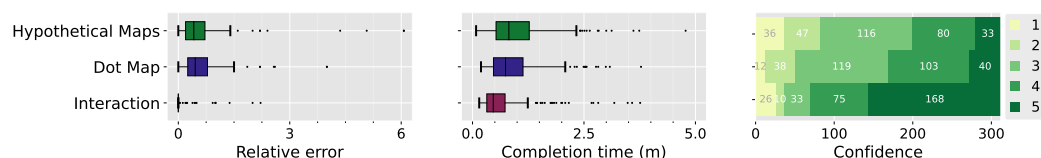

Figure 6. Results of the compare distributions task.

Estimate Mean. Since the DIM technique does not provide a specific functionality to find the mean value for a region, a larger relative error was expected, confirmed in the user study (Figure 7 (left)). The KW test showed a significant difference between all techniques, and the MWW test showed a significant difference between all pairs, except between the DIM and DDM groups. This supports H4. In terms of completion time, the DIM technique registered the largest completion time (Figure 7 (center)). This was already expected, given the need to interact with the slider multiple times to perform the estimation. The KW test showed a significant difference and the MWW test showed a significant difference between all techniques, except between the DDM and HOM. This reflects what was already hypothesized in H2, as the DIM is not the best proposal in terms of completion time.
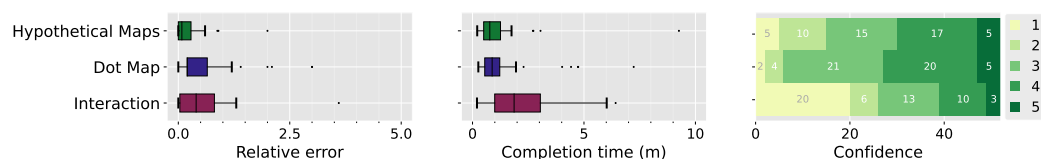

Figure 7. Results of the estimate mean task.

For the first time among all tasks, the DIM technique had the lowest confidence scores since the adapted visual metaphor does not directly allow one to find the results as in the

other tasks (Figure 7 (right)). The KW test showed that there is a significant difference between the techniques. The MWW test showed a significant difference, except between DDM and HOM.

4.4. Impact of Distance, Size and Variance on the Comparison Task

We now turn to the investigation of the impact of three different variables on the relative error and completion time of the visualization techniques tested: the distance between geographical regions (i.e., near and far), size of regions (i.e., small and large), and data variance in the regions (i.e., small and large), as shown in Figure 3. In each data set, we manually selected two regions that met the criteria of the variables.

Distance. In the DDM technique, the greater distance between geographical regions increased the relative error; in our user study, the median of the task that uses two distant areas is greater than the upper quartile of the short distance (Figure 8). This behavior is also similar among participants in the HOM group. In both cases, the difference between the variables was statistically significant according to the MWW test. The distance between areas did not impact the relative error in the DIM technique; both conditions had a small error, with a median of 0. The user study did not show any significant impact in terms of completion time or confidence in all three techniques.
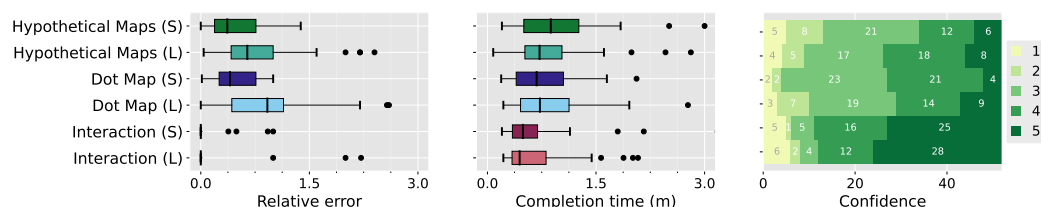

Figure 8. Results of the task measuring the impact of the distance between geographical regions.

Size. The size of the geographical regions (small and large) did not significantly impact the relative error or completion time of any of the studied techniques when we consider both data sets (Figure 9). However, size does play a role in the DDM technique when considering only the taxi data set. There were smaller errors when considering large geographical regions, with the median of this case below the lower quartile of the condition with small geographical regions, given that the larger space fits more points and thus better represents the distribution of the region. This result was statistically significant according to the MWW test.
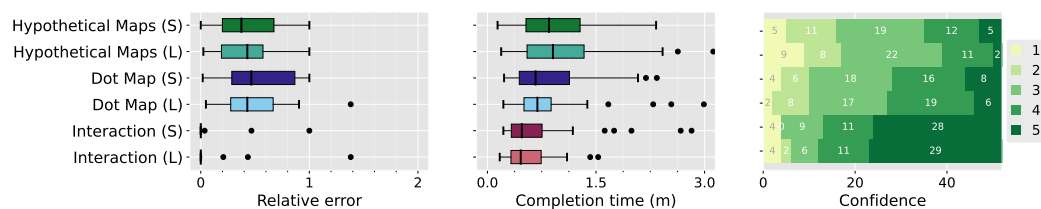

Figure 9. Results of the task measuring the impact of the size of the geographical regions.

Variance. When considering both data sets together, data variance did not have a significant impact on any of the techniques, when considering relative error or completion time (Figure 10). However, like the size variable, variance significantly impacted the DDM technique when considering the taxi data. There were smaller errors when considering the large variance case; the median relative error for the large variance condition is below the lower quartile of the small variance case.
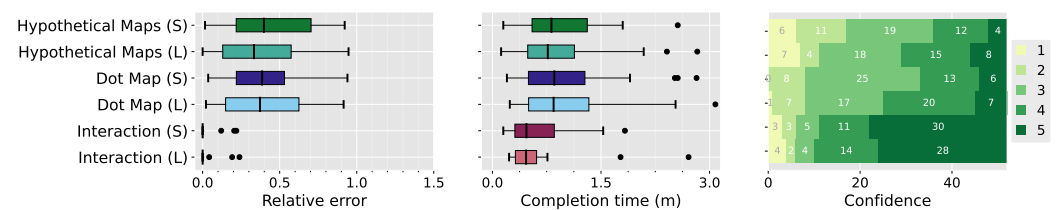
Figure 10. Results of the task measuring the impact of the data variance between geographical regions.

4.5. Comments and Feedback from Participants

In this section, we summarize the qualitative comments given by the participants both at the end of the experiment (fourth part of the study) and recorded during the sessions under the think-aloud protocol. These comments shed light on possible interpretation and usability issues that the users had while using the different visualization techniques.

Participants of the DIM technique had, in general, positive impressions. A participant of the DIM group mentioned that *"when you learn and understand the tool it becomes easier"*. While this was somewhat expected, we highlight that performing the tasks still required a lot of attention for interpreting the question correctly. This was the subject of the feedback given by a different participant: In fact, the impression was that the tool made some of the questions "too simple" to the point that they thought that there could be something tricky on them. This can be seen in the following comment: *"I was afraid of trusting the tool entirely. After you learn it, it is very easy to use. However, I still tried to spend time double-checking my answers"*. We believe that the fact that the DIM approach did not give the users access to the underlying data might be the reason for this insecurity. Finally, the task of estimating the mean was mentioned by several participants, with one saying *"I was just in doubt with the mean"*.

Regarding the DDM technique, many participants reported difficulty in performing the tasks and also the need to constantly use zoom and pan interactions. For example, one participant said that *"I found it hard to estimate the proportion of dots of the same color in an area, so I started counting. I had to zoom in the area to be able to count"*. Another interesting feedback given by the participants was that, as expected by our hypothesis H5, the distance between regions increases the difficulty in comparing the distributions. As stated by one participant: *"it (the estimation process) is time-consuming since there are cases where regions are far apart, and others are close together. In some cases, it is necessary to zoom in"*. This was also a source of insecurity, as stated in the following comment by a different participant: *"I had difficulty in analyzing (these visualizations) due to the counting process: distances make it more difficult, and it seems like an uncertain process. The tool should give us some sort of feedback to enable better analysis of the distribution"*. As other comments have shown, this view of an "uncertain process" might result from the fact that participants found it difficult to spot the differences in the color patterns.

The comments on the HOM technique showed the evidence of an intuitive fact: animation might make the estimation process difficult. Furthermore, as shown in previous works that used animation for geographical visualization, the large number of changing objects in the visualization can cause confusion: *"there are simply too many things to keep track of (...) I found this wearing"*. One important feature noticed while we observed the users was that sudden changes in the color of a (possibly distant) region from the user's current point of interest could catch the user's attention and make him or her lose focus on the intended analysis. Participants also reported not being confident using this visualization: *"I think this is a pretty nice visualization, but you get insecure, especially for someone like me used to calculations"*. Finally, as with the DDM, many subjects found it difficult to perform tasks for which the regions were far apart or with similar distributions: *"it was very difficult to compare regions that are very far from each other, or that the colors are similar"*. Another participant also mentioned the distance: *"for me the closer the distance, the easier to answer"*.

5. Discussion

The results of our user study confirm that the DIM technique was the one with higher accuracy and lower completion time for the retrieve value, find extremes, and comparison tasks. Participants also reported a much higher confidence score when compared with the other two techniques. This combination indicates that DIM yields high confidence coupled with good performance. This desirable property was named by Ferreira et al. [8] as *justified confidence*. This can be explained by the fact that the interaction allows for the exact quantification of the probability distribution, while dot map and hypothetical maps rely on counting or memorization of partial results. As pointed out by the participants, this creates a level of insecurity when answering the questions. Both HOM and DDM had large relative errors, especially for the retrieve value task, with medians larger than 0.25. Errors of this size are likely not to be suitable for real applications. The quantitative results support our initial hypotheses H1, H2, and H3.

On the other hand, some users were insecure in using the DIM approach when estimating means. We hypothesize that this is because this technique did not give them access to the underlying data, but only to the quantified probabilities. In this task, HOM had higher accuracy and lower completion time, while DIM presented the highest errors. Past work did not use interaction for this particular task [6,49], also indicating that one would need another approach to communicate the mean.

Finally, as expected, when performing comparisons, users were affected by the spatial distance factor in both DDM and HOM, which limits their use in these scenarios. While we did not detect any significant impact of region area and distribution variance, we intend to perform more detailed studies in this scenario in which we could vary different degrees of these factors and check their influence on the participants' performance.

6. Conclusions, Limitations and Future Work

In the present study, we compared three techniques that support probabilistic analytical tasks and, as a consequence, allow users to make proper use of the probability distributions in the analysis process. As discussed in the previous section, each one of the different techniques has negative and positive aspects.

While HOM and DDM yielded large relative errors, they have some desired features that are not present in DIM. First, we believe that, similar to the original version of (non-spatial) hypothetical outcome plots [16], HOM presents a natural way of depicting uncertainty in the sense that users naturally interpret this animation as an uncertain display. Furthermore, although not directly tested in our study, we also hypothesize that this is a more suitable metaphor for dealing with non-expert users. On the other hand, we believe that the DIM approach is more useful for an exploration scenario done by expert users that know the exact meaning of the different interactive tools provided.

Concerning DDM, one important particular feature is that, among the techniques presented in this paper, it is the only one that can perform *visual spatial aggregation*. More clearly, performing one of the analytical tasks for a macro-region corresponding to the union of two or more regions on the map is done by estimating the density of points as if the macro-region was one of the original subdivisions on the map. This feature is important, since geographical visualization often involves analyzing patterns in different scales. However, one negative feature of DDM is that when using it for larger spatial regions, users can be led to interpret the location of individual dots as actual information provided by the map. While we acknowledge that this might be an actual problem, we believe that properly setting the dot size and placement strategy can create layouts that are both regular and non-natural, which indicate to the user that these are not actual locations. In our current paper, we selected the dot size as a good compromise between showing the distribution details and being visible. However, how to automatically set the dot size and dot placement, to optimize users' accuracy, given the region geometries and the distributions is an interesting direction of future work. While this technique generally attained the worst results in our study, we believe that it has a great deal of

potential for future developments. One interesting direction of research is to investigate whether ordering the dots spatially with respect to their sample value can improve the users' accuracy when performing the tasks discussed.

Finally, one important feature of HOM is that among the three techniques present, it is the only one that can visually represent correlation among the different regions. While we acknowledge that estimating the correlation between two distributions is an important analytical task itself, we chose not to include it in our current study given that only one technique could represent it. A recent study by Peña-Araya et al. [61] explores the design of static glyph representations with the goal of estimating spatio-temporal correlations. An interesting direction for future work is to compare the glyphs proposed in this related work with HOM for this particular task.

We now discuss some limitations of our study. Even though we used two data sets with very different numbers of regions, the influence of this factor was not clear. We plan to study this influence and also the impact of factors such as the number and shape of geographical regions on the user's performance. Our user study did not consider different sampling processes for DDM, as well as the impact of dot size. We plan to evaluate this in a future study, taking into account different sorting options (e.g., sorted by value left to right) as well as the relationship between these options and the size of geographical regions. We also recognize the fact that our user study makes use of a limited set of geographical regions. In future work, we plan to study how well the results generalize to a wider range of geographic data sets. We also plan to evaluate other visualization approaches, such as bivariate, cartograms and dasymetric maps.

An important aspect of spatial data is that patterns can occur in different spatial resolutions. Therefore, another interesting direction for future work is to investigate the effectiveness of the discussed visual metaphors in performing visual spatial aggregation [61]. Finally, previous work presented evidence that better quantification leads to better decision-making. While intuitive, this has not been investigated in geographical data exploration scenarios. We intend to validate this in a future study.

Author Contributions: Conceptualization, N.F. and F.M.; methodology, S.S, C.V., E.S., M.L., N.F. and F.M.; software, S.S., C.V. and E.S.; formal analysis, S.S, C.V., E.S., N.F. and F.M.; data curation, E.S.; writing—original draft preparation, S.S., C.V., E.S.; writing—review and editing, M.L., N.F. and F.M.; visualization, S.S, E.S., N.F. and F.M.; supervision, N.F. and F.M. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Universidade Federal de Pernambuco (protocol code 23754819.5.0000.5208).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dasgupta, A.; Poco, J.; Bertini, E.; Silva, C.T. Reducing the analytical bottleneck for domain scientists: Lessons from a climate data visualization case study. *Comput. Sci. Eng.* **2016**, *18*, 92–100. [CrossRef]
2. Corral, Á.; González, Á. Power law size distributions in geoscience revisited. *Earth Space Sci.* **2019**, *6*, 673–697. [CrossRef]
3. de Souza, C.V.F.; da Cunha Luz Barcellos, P.; Crissaff, L.; Cataldi, M.; Miranda, F.; Lage, M. Visualizing simulation ensembles of extreme weather events. *Comput. Graph.* **2022**, *104*, 162–172. [CrossRef]
4. Wang, Z.; Ferreira, N.; Wei, Y.; Bhaskar, A.S.; Scheidegger, C. Gaussian cubes: Real-time modeling for visual exploration of large multidimensional datasets. *IEEE Trans. Vis. Comput. Graph.* **2016**, *23*, 681–690. [CrossRef] [PubMed]

5.　Moritz, D.; Fisher, D.; Ding, B.; Wang, C. Trust, but verify: Optimistic visualizations of approximate queries for exploring big data. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, 6–11 May 2017; ACM: New York, NY, USA, 2017; pp. 2904–2915.

6.　Pahins, C.A.L.; Ferreira, N.; Comba, J.L. Real-Time Exploration of Large Spatiotemporal Datasets Based on Order Statistics. *IEEE Trans. Vis. Comput. Graph.* **2020**, *26*, 3314–3326. [CrossRef] [PubMed]

7.　Matejka, J.; Fitzmaurice, G. Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics Through Simulated Annealing. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, 6–11 May 2017; ACM: New York, NY, USA, 2017; pp. 1290–1294.

8.　Ferreira, N.; Fisher, D.; Konig, A.C. Sample-Oriented Task-Driven Visualizations: Allowing Users to Make Better, More Confident Decisions. In Proceedings of the 2014 CHI Conference on Human Factors in Computing Systems, Toronto, ON, Canada, 26 April–1 May 2014; ACM: New York, NY, USA, 2014; pp. 571–580.

9.　Belia, S.; Fidler, F.; Williams, J.; Cumming, G. Researchers misunderstand confidence intervals and standard error bars. *Psychol. Methods* **2005**, *10*, 389. [CrossRef] [PubMed]

10.　Cumming, G.; Finch, S. Inference by eye: Confidence intervals and how to read pictures of data. *Am. Psychol.* **2005**, *60*, 170. [CrossRef]

11.　Correll, M.; Gleicher, M. Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE Trans. Vis. Comput. Graph.* **2014**, *20*, 2142–2151. [CrossRef]

12.　MacEachren, A.M.; Robinson, A.; Hopper, S.; Gardner, S.; Murray, R.; Gahegan, M.; Hetzler, E. Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know. *Cartogr. Geogr. Inf. Sci.* **2005**, *32*, 139–160. [CrossRef]

13.　Ślusarski, M.; Jurkiewicz, M. Visualisation of Spatial Data Uncertainty. A Case Study of a Database of Topographic Objects. *ISPRS Int. J. Geo-Inf.* **2019**, *9*, 16. [CrossRef]

14.　Kinkeldey, C.; MacEachren, A.M.; Schiewe, J. How to Assess Visual Communication of Uncertainty? A Systematic Review of Geospatial Uncertainty Visualisation User Studies. *Cartogr. J.* **2014**, *51*, 372–386. [CrossRef]

15.　Fernandes, M.; Walls, L.; Munson, S.; Hullman, J.; Kay, M. Uncertainty displays using quantile dotplots or cdfs improve transit decision-making. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; ACM: New York, NY, USA, 2018; pp. 1–12.

16.　Hullman, J.; Resnick, P.; Adar, E. Hypothetical outcome plots outperform error bars and violin plots for inferences about reliability of variable ordering. *PLoS ONE* **2015**, *10*, e0142444. [CrossRef] [PubMed]

17.　Kay, M.; Kola, T.; Hullman, J.R.; Munson, S.A. 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, San Jose, CA, USA, 7–12 May 2016; ACM: New York, NY, USA, 2016; pp. 5092–5103.

18.　Crampton, J. GIS and Geographic Governance: Reconstructing the Choropleth Map. *Cartogr. Int. J. Geogr. Inf. Geovisualization* **2004**, *39*, 41–53. [CrossRef]

19.　Adams, A.; Li, W.; Zhang, C.; Chen, X. The disguised pandemic: The importance of data normalization in COVID-19 web mapping. *Public Health* **2020**, *183*, 36. [CrossRef]

20.　Zhang, Y.; Sun, Y.; Padilla, L.; Barua, S.; Bertini, E.; Parker, A.G. Mapping the Landscape of COVID-19 Crisis Visualizations. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan, 8–13 May 2021; ACM: New York, NY, USA, 2021; pp. 1–23.

21.　Srabanti, S.; Marai, G.E.; Miranda, F. COVID-19 EnsembleVis: Visual Analysis of County-Level Ensemble Forecast Models. In Proceedings of the 2021 IEEE Workshop on Visual Analytics in Healthcare (VAHC), New Orleans, LA, USA, 24–25 October 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–5.

22.　Tukey, J.W. Mathematics and the picturing of data. In Proceedings of the International Congress of Mathematicians, Vancouver, BC, Canada, 1975; Volume 2, pp. 523–531.

23.　Potter, K.; Gerber, S.; Anderson, E.W. Visualization of uncertainty without a mean. *IEEE Comp. Graph. Appli.* **2013**, *33*, 75–79. [CrossRef]

24.　Wickham, H.; Stryjewski, L. 40 Years of Boxplots. Technical Report. 2011. Available online: https://vita.had.co.nz/papers/boxplots.html (accessed on 1 June 2022).

25.　Kale, A.; Kay, M.; Hullman, J. Decision-making under uncertainty in research synthesis: Designing for the garden of forking paths. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019; ACM: New York, NY, USA, 2019; pp. 1–14.

26.　Blumenschein, M.; Debbeler, L.J.; Lages, N.C.; Renner, B.; Keim, D.A.; El-Assady, M. v-plots: Designing Hybrid Charts for the Comparative Analysis of Data Distributions. *Comput. Graph. Forum* **2020**, *39*, 565–577. [CrossRef]

27.　Correll, M.; Gieicher, M. Implicit uncertainty visualization: Aligning perception and statistics. In Proceedings of the Visualization for Decision Making under Uncertainty Workshop, Chicago, IL USA, 26 October 2015.

28.　Nguyen, F.; Qiao, X.; Heer, J.; Hullman, J. Exploring the Effects of Aggregation Choices on Untrained Visualization Users' Generalizations From Data. *Comput. Graph. Forum* **2020**, *39*, 33–48. [CrossRef]

29.　Hullman, J. Why Authors Don't Visualize Uncertainty. *IEEE Trans. Vis. Comput. Graph.* **2019**, *26*, 130–139. [CrossRef]

30. MacEachren, A.M. Visualizing uncertain information. *Cartogr. Perspect.* **1992**, *13*, 10–19. [CrossRef]
31. Potter, K.; Wilson, A.; Bremer, P.T.; Williams, D.; Doutriaux, C.; Pascucci, V.; Johnson, C.R. Ensemble-vis: A framework for the statistical visualization of ensemble data. In Proceedings of the 2009 IEEE International Conference on Data Mining Workshops, Miami, FL, USA, 6 December 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 233–240.
32. MacEachren, A.M.; Brewer, C.A.; Pickle, L.W. Visualizing georeferenced data: Representing reliability of health statistics. *Environ. Plan.* **1998**, *30*, 1547–1561. [CrossRef]
33. Kinkeldey, C.; Mason, J.; Klippel, A.; Schiewe, J. Evaluation of noise annotation lines: Using noise to represent thematic uncertainty in maps. *Cartogr. Geogr. Inf. Sci.* **2014**, *41*, 430–439. [CrossRef]
34. Kubíček, P.; Šašinka, Č. Thematic uncertainty visualization usability–comparison of basic methods. *Ann. GIS* **2011**, *17*, 253–263. [CrossRef]
35. Correll, M.; Moritz, D.; Heer, J. Value-suppressing uncertainty palettes. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal QC Canada, 21–26 April 2018; ACM: New York, NY, USA, 2018; pp. 1–11.
36. Edwards, L.D.; Nelson, E.S. Visualizing data certainty: A case study using graduated circle maps. *Cartogr. Perspect.* **2001**, *38*, 19–36. [CrossRef]
37. Sanyal, J.; Zhang, S.; Dyer, J.; Mercer, A.; Amburn, P.; Moorhead, R. Noodles: A tool for visualization of numerical weather model ensemble uncertainty. *IEEE Trans. Vis. Comput. Graph.* **2010**, *16*, 1421–1430. [CrossRef]
38. Roth, R.E. Dot density maps. *Encycl. Geogr.* **2010**, 787–790.
39. Dmowska, A.; Stepinski, T.F. Racial Dot Maps Based on Dasymetrically Modeled Gridded Population Data. *Soc. Sci.* **2019**, *8*, 157. [CrossRef]
40. Chua, A.; Moere, A.V. Binsq: Visualizing geographic dot density patterns with gridded maps. *Cartogr. Geogr. Inf. Sci.* **2017**, *44*, 390–409. [CrossRef]
41. Hey, A. Automated dot mapping: How to dot the dot map. *Cartogr. Geogr. Inf. Sci.* **2012**, *39*, 17–29. [CrossRef]
42. Evans, B.J. Dynamic display of spatial data-reliability: Does it benefit the map user? *Comput. Geosci.* **1997**, *23*, 409–422. [CrossRef]
43. Kardos, J.; Benwell, G.; Moore, A. The visualisation of uncertainty for spatially referenced census data using hierarchical tessellations. *Trans. GIS* **2005**, *9*, 19–34. [CrossRef]
44. Ehlschlaeger, C.R.; Shortridge, A.M.; Goodchild, M.F. Visualizing spatial data uncertainty using animation. *Comput. Geosci.* **1997**, *23*, 387–395. [CrossRef]
45. Cybulski, P. An Empirical Study on the Effects of Temporal Trends in Spatial Patterns on Animated Choropleth Maps. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 273. [CrossRef]
46. Howard, D.; MacEachren, A.M. Interface design for geographic visualization: Tools for representing reliability. *Cartogr. Geogr. Inf. Syst.* **1996**, *23*, 59–77. [CrossRef]
47. Faïz, S.; Boursier, P. Geographic data quality: From assessment to exploitation. *Cartogr. Int. J. Geogr. Inf. Geovisualization* **1996**, *33*, 33–40. [CrossRef]
48. Cliburn, D.C.; Feddema, J.J.; Miller, J.R.; Slocum, T.A. Design and evaluation of a decision support system in a water balance application. *Comput. Graph.* **2002**, *26*, 931–949. [CrossRef]
49. Alberti, K. Web-Based Visualization of Uncertain Spatio-Temporal Data. Master's Thesis, Utrecht University, Utrecht, The Netherlands, 2013.
50. Elmqvist, N.; Fekete, J.D. Hierarchical aggregation for information visualization: Overview, techniques, and design guidelines. *IEEE Trans. Vis. Comput. Graph.* **2009**, *16*, 439–454. [CrossRef]
51. Ferreira, N.; Poco, J.; Vo, H.T.; Freire, J.; Silva, C.T. Visual exploration of big spatio-temporal urban data: A study of New York City taxi trips. *IEEE Trans. Vis. Comput. Graph.* **2013**, *19*, 2149–2158. [CrossRef]
52. Lins, L.; Klosowski, J.T.; Scheidegger, C. Nanocubes for Real-Time Exploration of Spatiotemporal Datasets. *IEEE Trans. Vis. Comput. Graph.* **2013**, *19*, 2456–2465. [CrossRef]
53. Doraiswamy, H.; Tzirita Zacharatou, E.; Miranda, F.; Lage, M.; Ailamaki, A.; Silva, C.T.; Freire, J. Interactive Visual Exploration of Spatio-Temporal Urban Data Sets Using Urbane. In Proceedings of the 2018 International Conference on Management of Data, SIGMOD '18, Houston, TX, USA, 10–15 June 2018; ACM: New York, NY, USA, 2018; pp. 1693–1696.
54. Miranda, F.; Lins, L.; Klosowski, J.T.; Silva, C.T. TopKube: A Rank-Aware Data Cube for Real-Time Exploration of Spatiotemporal Data. *IEEE Trans. Vis. Comput. Graph.* **2018**, *24*, 1394–1407. [CrossRef]
55. Andrienko, N.; Andrienko, G. *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*; Springer: Berlin, Heidelberg, Germany, 2006.
56. Amar, R.; Eagan, J.; Stasko, J. Low-level components of analytic activity in information visualization. In Proceedings of the IEEE Symposium on Information Visualization, Minneapolis, MN, USA, 23–25 October 2005; IEEE: Piscataway, NJ, USA, 2005.
57. Kick the bar chart habit. *Nat. Methods* **2014**, *11*, 113. [CrossRef]
58. Weissgerber, T.L.; Milic, N.M.; Winham, S.J.; Garovic, V.D. Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. *PLoS Biol.* **2015**, *13*, e1002128. [CrossRef] [PubMed]
59. Rousselet, G.A.; Foxe, J.J.; Bolam, J.P. A Few Simple Steps to Improve the Description of Group Results in Neuroscience. *Eur. J. Neurosci.* **2016**, *44*, 2647–2651. [CrossRef] [PubMed]

60. Lucchesi, L.R.; Wikle, C.K. Visualizing uncertainty in areal data with bivariate choropleth maps, map pixelation and glyph rotation. *Stat* **2017**, *6*, 292–302. [CrossRef]
61. Peña-Araya, V.; Pietriga, E.; Bezerianos, A. A Comparison of Visualizations for Identifying Correlation over Space and Time. *IEEE Trans. Vis. Comput. Graph.* **2019**, *26*, 375–385. [CrossRef] [PubMed]