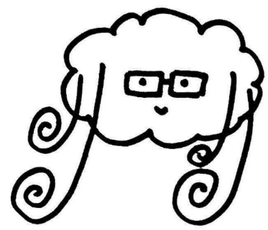
Helping Users Make Decisions by Visualizing Weather Uncertainty

Jyalu Wu, Andrew Nolan, Matthew St Louis, Imogen Cleaver-Stigum

Jyalu
(snow)

Andrew
(thunderstorm)



Matt
(wind)

Imogen
(tornado)

Table of Contents

1 Overview and Motivation

This project aims to determine the best way to visualize uncertainty in weather predictions. Our group has spent a lot of time looking into uncertainty visualizations, and we feel the final project may be one place we can apply them in an interesting way. Specifically, we are inspired by quantile dot plots and hypothetical outcome plots.

We see precipitation uncertainty as a concise metric for binary decision-making. We presented participants with several visualizations that display the uncertainty around precipitation and ask them how they would handle certain situations (e.g. paying a small fee to move a community event to an indoor space, going for a walk outside, rescheduling a hike with friends). We also asked participants to estimate the expected level of precipitation at a certain time based on the visualizations. Our goal was to find out how the visualization type affects the participant's answers, what visualization types make the user feel most confident in their decisions, and what visualization types are most accurately readable to participants.

2 Related Work

There are many publications we can draw from and build upon for our project, such as:

- **When (ish) is my bus**: attempts to visualize uncertainty in bus arrival times for decision-making. The authors of this study sought out to improve the existing OneBusAway app by visualizing uncertainty. They surveyed users on what aspects of bus arrival times were useful and how they used the app in their everyday lives. From this, they designed a few different uncertainty visualizations, including the novel quantile dot plot. This is very similar to the research we plan to conduct. We would likely also start by assessing what weather data people need to see and ideating potential representations of it. We would then ideate on what visualizations might best meet the needs that users express or that we find in our research. In addition to mimicking this process, quantile dot plots may be an interesting visualization to examine.
- **Hypothetical Outcome Plots**: This research proposed a novel uncertainty visualization that represents an uncertain distribution dynamically rather than statically. The idea is to have a chart that periodically updates to show a new hypothetical draw from an uncertainty distribution, and the user can mentally aggregate these draws to get a sense of the overall distribution. We used this technique to visualize uncertainty in weather data dynamically.
- **Value-Suppressing Uncertainty Pallets**: This research focuses on examining color pallets that can communicate uncertainty efficiently and effectively. These color pallets allocate more colors to more certain data and fewer colors to less certain data, which can help clear visual noise from a map. We are interested in attempting to apply this technique to a map showing expected precipitation or some other metric.
- **Sanyal et al.** created **Noodles**: a tool for visualization of numerical weather model ensemble uncertainty. This tool visualizes ensemble uncertainty for water-vapor mixing ratio, perturbation potential temperature, and perturbation pressure. They quantified uncertainty using individual

ensemble member standard deviation, interquartile range, and the width of the 95% confidence intervals after performing bootstrap aggregation to ensure the data followed a normal distribution. Using *uncertainty glyphs*, *uncertainty ribbons*, and *Spaghetti plots* they designed a user interface for exploring weather uncertainty. This research was evaluated by two meteorologists. The tool appeared to be effective for seeing areas of uncertainty in the weather. However, the features used and the target audience of this tool are for expert meteorologists. Our team can build upon the visualizations used in this research but with specifically precipitation data that can be easily interpreted and used by casual users interested in the weather.

- **The Effect of Uncertainty Visualizations on Decision Making in Weather Forecasting**: This paper examines a few different ways of visualizing weather uncertainty in wind speeds. The technique it proposes (using a combination of maps for the worst case, average case, and variability, each with its own hue-based color scale) does not seem particularly useful. This paper's main use for us lies in understanding what to avoid when designing our own visualizations.
- **Visual Reasoning Strategies and Satisficing: How Uncertainty Visualization Design Impacts Effect Size Judgments and Decisions**: This paper looks at how people perceive uncertainty based on how far apart visually two plots are. It also looks at whether this changes when the mean is superimposed over the uncertainty plot (for example, a normal distribution bell curve with the mean highlighted) and finds that including/emphasizing the mean makes people less likely to accurately perceive the uncertainty. However, the paper also finds that accuracy of the uncertainty perceptions does not necessarily benefit the viewers' decision making. This is relevant to us because we want to understand which types of plots are best for visualizing uncertainty, as well as how people are going to use this uncertainty in their everyday decision making about precipitation.
- **In pursuit of error: A survey of uncertainty visualization evaluation**: In this paper, **Jessica Hullman et. al.** designed a *taxonomy* to help improve the process of evaluating the effectiveness of an uncertainty visualization. This taxonomy has six levels of decisions that are used to evaluate vizzes, and depending on which decision you decide on for each level, your *evaluation path* will differ. They also give recommendations for evaluators of uncertainty vizzes. We will try to follow both the taxonomy and the recommendations when self-evaluating our vizzes throughout the project, from building them to analyzing survey responses.
- **Why Authors Don't Visualize Uncertainty**: This paper outlines the reasons why visualization authors choose not to visualize uncertainty after surveying 90 of them. The paper also references the contradiction that many authors state that uncertainty visualization is extremely important but choose not to delve into that field. This paper will allow us to understand the misconceptions and general attitudes towards uncertainty visualizations, as well as understanding some problems we might run into while building them.

3 Questions

What questions are you trying to answer? How did these questions evolve over the course of the project? What new questions did you consider in the course of your analysis?

We started this project to learn more about visualization research in general and to learn more about what uncertainty visualizations are effective within the specific problem domain of weather uncertainty. Our initial questions can be broken into the following groups:

- Questions about visualization research
  - What steps go into researching and preparing data to visualize?
  - What important axes do visualization researchers often consider in an experiment like this? (e.g. how much effort participants need to expend to read the visualization, how informative the visualizations are, how much participants like the visualizations)
- Questions about visualizing weather uncertainty
  - Would weather app users be interested in being able to visualize the uncertainty of precipitation?
  - Can we do better than the bar charts or other visualizations that some weather apps currently use?

As we prepared for the project through research, we answered some of our preliminary questions. We found common axes to compare uncertainty visualizations, and we found some best practices, such as allowing respondents to communicate that the visualization is too unclear to make a judgement call on. We answered our questions about data generation and experiment setup as we put together our experiment. Our experimental questions are addressed in the Evaluation section.

4 Data

We decided to generate the precipitation prediction data artificially because historical data for weather predictions is rare to find, especially historical data that would include the predicted chance of rain for the future. Our artificially generated data included the predicted amount of rain in inches per hour as well as the predicted chance of precipitation (0%-100%) for that hour. All of the values are rounded to the nearest .05. We used references of real weather data for Worcester to make our data realistic enough for our purposes. The code to generate this data can be found in the repo in the file: react-firebase/src/data-generation.js.

4.1 Probability of Precipitation

We generate the data in batches of 24 (or another number of) hours at a time. The probability of precipitation (POP) ranges from 0 to 1 and is determined randomly with some conditions. The first data point is a completely random number. To generate the subsequent POP values for all the remaining 23 hours of the day, we first determine randomly whether the values will be ascending or descending. We then take a value within 25% of the previous value, with a 80% chance of going in the specified direction (ascending/descending). For each subsequent data point, we use this method to determine the value, if the value is above the upper end of the range [0,1], it is capped at the upper limit. The same goes for the lower limit of the range. In addition, if the values reach the upper or lower limits, we switch the direction (ascending/descending) of the POP values.

We based this method off logic about the way the POP behaves in real life, looking at real weather data from various sources like OpenWeatherAPI [10] and other weather prediction sources, as well as sites that explain meteorological methods like NOAA [9]. We do this because the POP does not tend to change by huge values, like 90-100%, within a single hour. In addition, the POP does not tend to jump up and down repeatedly within a span of just a few hours (10%, 55%, 5%, 60%, 10%, 70% is a relatively improbable sequence for hourly data). We acknowledge that we have simplified the weather data with these methods in some ways. For example, some scenarios that are unlikely but possible in real data are completely impossible in our generated data. It is also simplified in that all of the random number generation is getting random numbers uniformly distributed over the given range, rather than using any other sort of distribution that might be more realistic.

Our POP data generation methods are also very unlikely to produce completely sunny days. This does make our data less realistic. However, it is appropriate for our project that our visualizations should show disproportionately days that do have some chance of precipitation so that the experiment participants can make interesting, nontrivial judgments about the precipitation.

4.2 Amount of Precipitation

The amounts of precipitation (in inches) are generated based on the POP values for each hour. We essentially take a random number between 0 and .6 inches/hour and multiply it by the POP value for that hour. We do this because in real weather data that we have observed, there is generally a correlation between the POP and the expected quantity of rain. For example, if there is more than .5 inches per hour of precipitation expected, the POP is almost always 1 in the real data we observed. If the POP is .1 or lower, the amount of rain is generally very small. We based the range of 0 to .6 inches per hour on the meteorological definitions of precipitation categories. We wanted our range of precipitation predictions to be able to cover all of the categories:

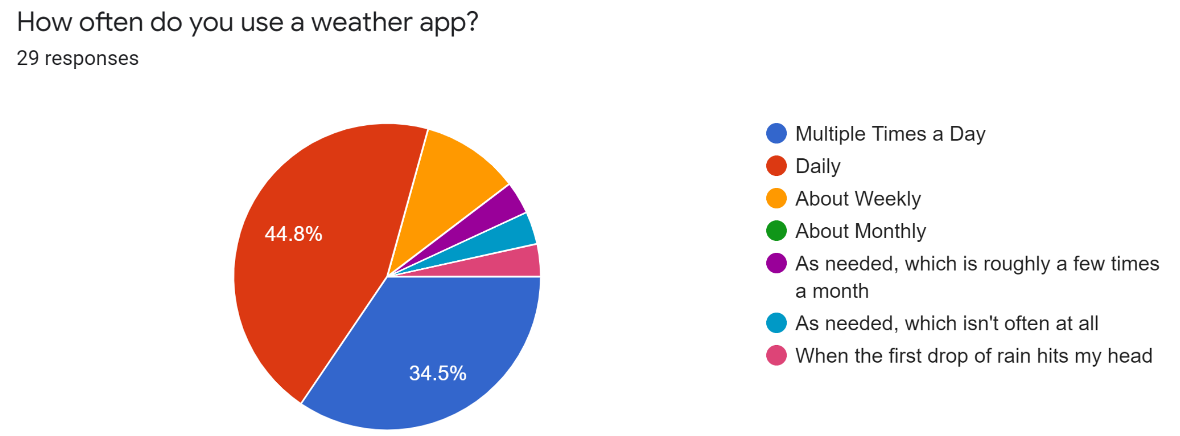
- .01 inches per hour is the smallest amount that the meteorological equipment is able to detect, so this is the lowest value categorized as “light precipitation” (“drizzle” or ‘flurry”, for example)
- .01-.1 inches per hour is “light precipitation”
- .1-.3 inches per hour is “moderate precipitation”
- .3+ inches per hour is “heavy precipitation” [9]

5 Preliminary Survey Results

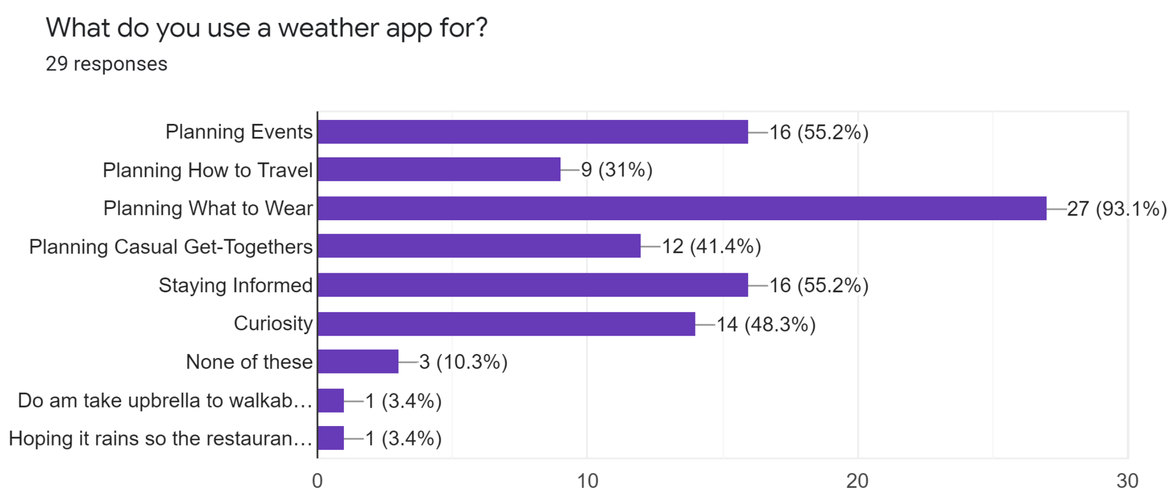
Before creating the experiment to test our visualizations, we sent out a survey to get preliminary results about how people use weather apps. We gathered responses from 29 participants.

5.1 General Use of Weather Apps

We asked participants about their general use of weather apps. We found that about 80% of participants use a weather app at least daily.

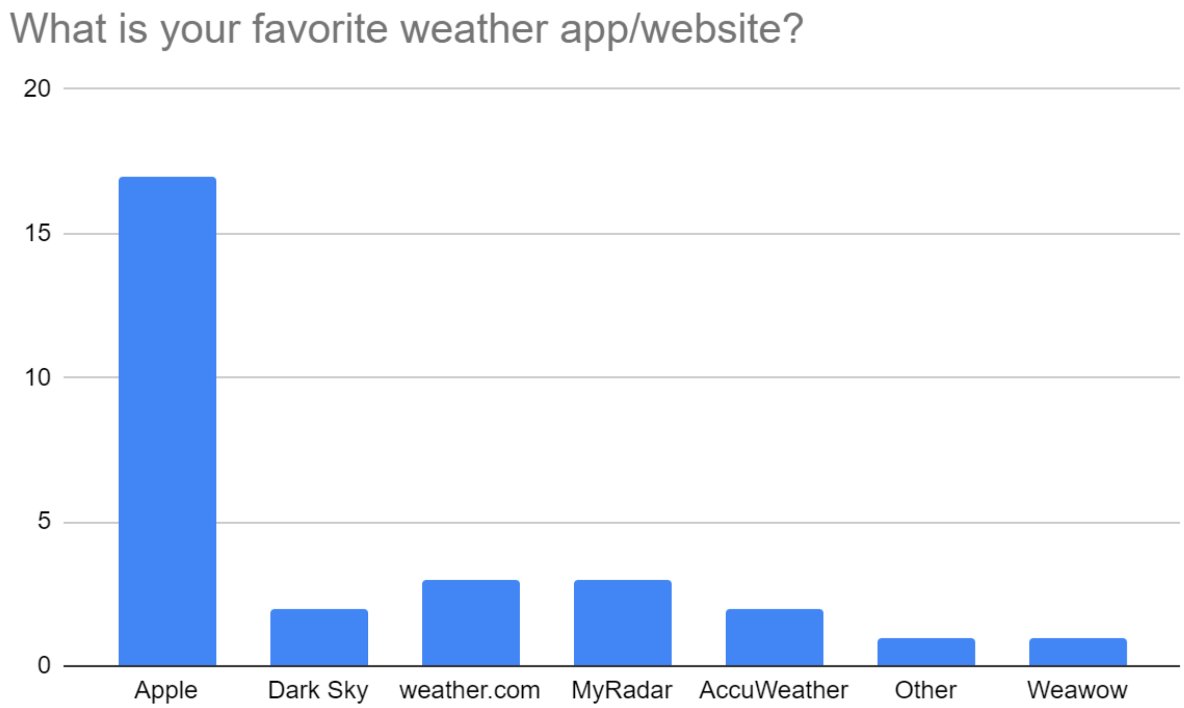


We also asked participants what they use their weather app for. Almost all participants use the weather app for planning what to wear, and about half also use it for planning events, staying informed, and curiosity.



5.2 Preferences about Weather Apps

Next, we wanted to find out about the participants' preferences about weather apps. We asked participants about their favorite weather apps and found that most participants like the default weather app on the iPhone (Apple).

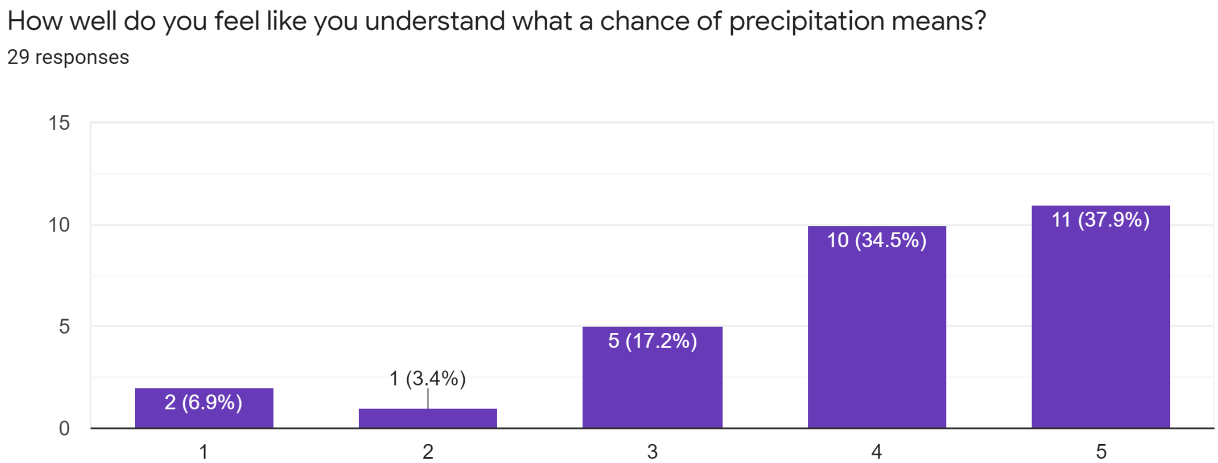


We also asked participants what they like and dislike about their weather apps. The participants like weather apps that provide data at small time intervals, such as hourly data, or even data about every 5 minutes. They also appreciated convenience, simplicity, and being quick to use and read.

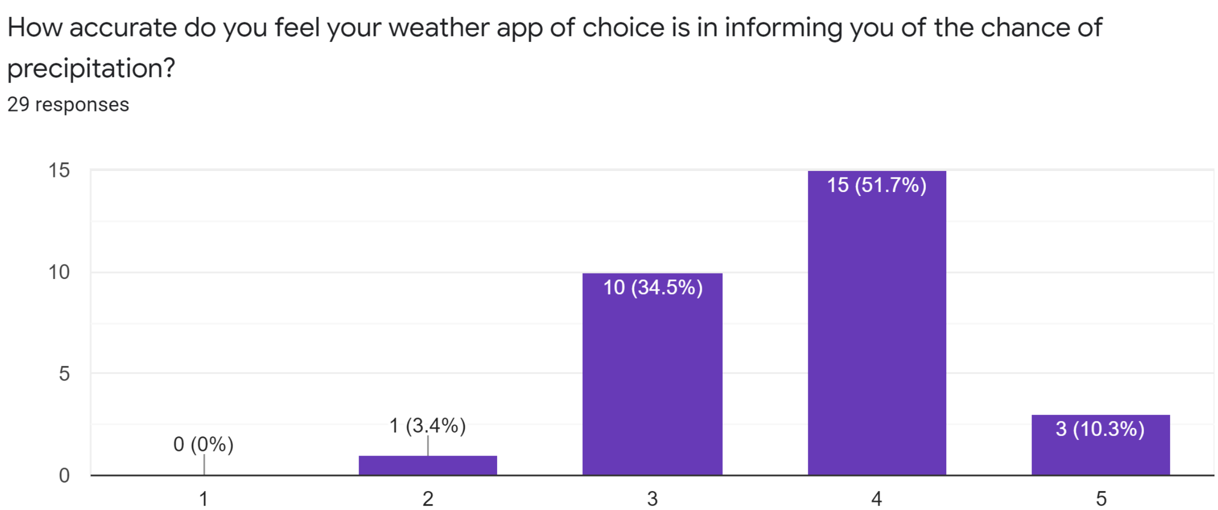
Participants were dissatisfied that their weather apps were not always accurate enough or precise enough. Several participants also wanted a radar/map view. However, several participants also said that they had never considered what more they might want in a weather app and did not have any ideas.

5.3 Understanding of Weather Vis

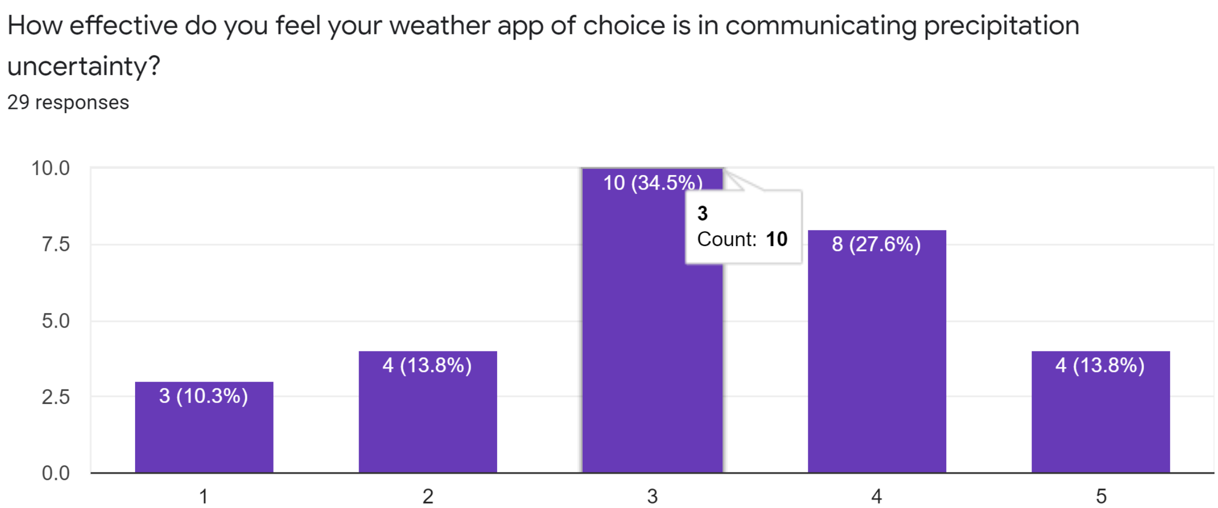
We also wanted to know how well participants understand their weather app and how well the weather app informs their decisions. We found that the majority of participants rated themselves as understanding what a chance of precipitation means extremely well (5) or very well (4).



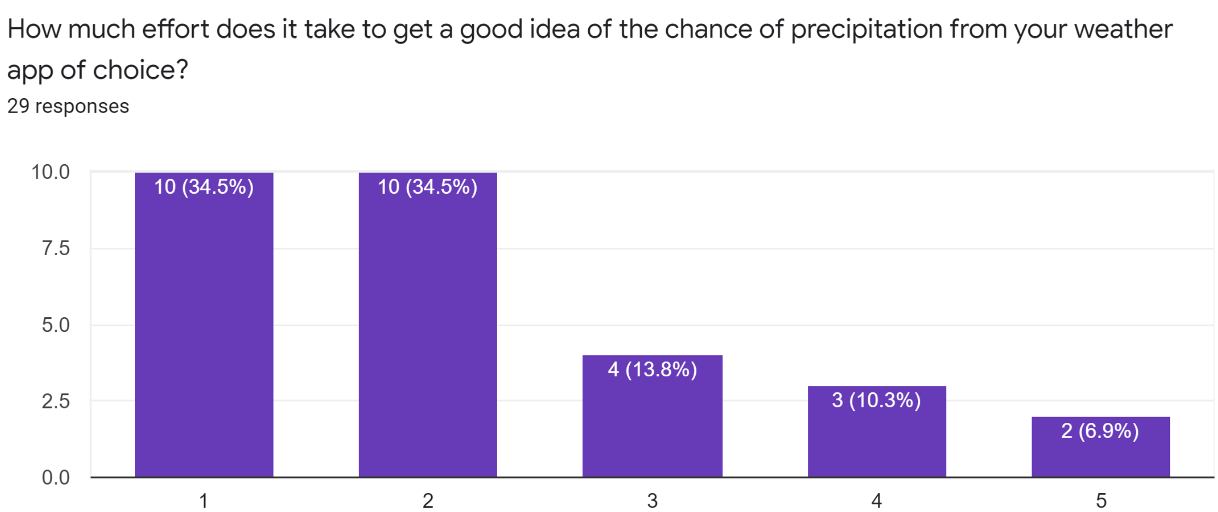
Most participants feel that the weather app is moderately accurate or very accurate in informing them about the chance of precipitation:



Participants overall felt that the weather app does a moderately effective job of communicating precipitation uncertainty:

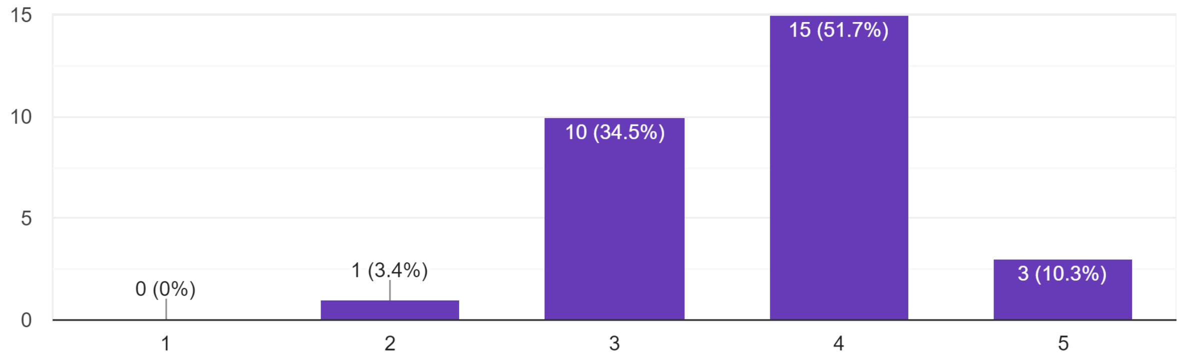


The majority of participants felt that they needed low effort (1 or 2) to get a good idea of the chance of precipitation from their weather apps:



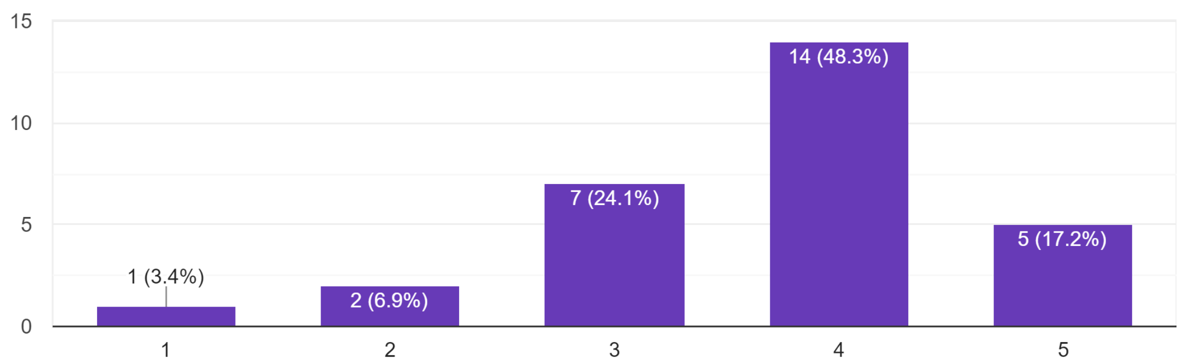
Most participants felt that their weather apps are very reliable (4) for making decisions based on precipitation:

How accurate do you feel your weather app of choice is in informing you of the chance of precipitation?
29 responses



Finally, most participants said that they would be likely or extremely likely (4 or 5) to use a visualization of weather uncertainty in their weather apps:

If a visualization of precipitation uncertainty was added to your favorite weather app, how likely would you be to use it?
29 responses



Overall, this survey has shown that most people are reliant on weather apps for daily decision making. It also showed that while weather apps do a fairly good job of communicating precipitation uncertainty, there is room for improvement.

5.4 Comparisons of Weather Vizzes

We also asked survey participants to describe which of the weather visualizations shown in Appendix 1 they preferred and why. Overall, the users preferred the visualizations that were simple and easy to read. However, the apps that were too simple were not preferable because they did not provide enough information, like hour-by-hour or even more frequent forecases. The favorite app of the group was Apple Weather. Participants liked that it had very specific precipitation data at small time intervals, good color coding and pictures, and a pleasant aesthetic appearance.

6 Design Evolution

What are the different visualizations you considered? Justify the design decisions you made using the perceptual and design principles you learned in the course. Did you deviate from your proposal?

Talk about the papers we read for inspiration

6.1 Designs Considered

We considered several different vizzes for weather uncertainty before deciding on the ones to include in our experiment. First, we considered some types of map-based vizzes. This included storm path visualizations to show where the precipitation was travelling over time, as well as maps using value suppressing uncertainty palettes to communicate the amount and certainty of precipitation for each area.

We also considered several types of bar charts, including plain static bar charts for a baseline, hypothetical outcome plots (HOPs), and bar charts with error bars corresponding to the certainty of precipitation. The bar charts with errors bars would have had the bar heights corresponding to the amount of rain and the error bars scaled to show the POP.

We also considered quantile dot plots, as well as text for a baseline to compare the vizzes against.

6.2 Designs Chosen and Revisions

We decided to implement three vizzes: static bar charts, hypothetical outcome plots, and quantile dot plots. We also decided to implement the text representation of the forecast as a baseline. For each of our vizzes, we display the precipitation forecast (amount and probability of precipitation) over the hours of the day.

The static bar charts do not convey uncertainty - they are also a baseline. The heights correspond to the amount of rain expected each hour. These plots were simple to implement so we did not end up having to make many revisions to our design.

The HOPs were designed similarly to the static bar charts. As we were implementing them, we considered several different ways of controlling the distributions over which the bars would animate. For example, some implementations of HOPs would have the value for a bar change with a uniform distribution in a constant interval around the true value for that bar. This approach did not correspond to how we wanted to convey the different values of uncertainty based on the POP values. Since the POP values are the probability that there will be any precipitation at all, we decided to make there be a 1-p chance of the bar having a height of 0. For example, if there was an 80% chance of rain, the bar would have height 0 20% of the time. (1-p)*2/3 of the time, the bar was within about 20% of the true value for the predicted amount of precipitation, within which we used a uniform distribution (random numbers). (1-p)/3 of the time, the bar was within 85% of the predicted amount and 0. For example, if there was a 90% chance of .5 inches of rain, the bar would be at height 0 10% of the time, between .4 and .6 inches 60% of the time, and between heights 0 and 0.4 30% of the time. We decided on this distribution based on trial and error, the expectations of what actual weather data would look like and what the POP metrics mean, and simplicity of implementation.

For the quantile dot plots (QDPs), we had to change our design significantly over the course of the implementation. Initially we were imagining only using one QDP, but we realized that we would have to animate the QDPs to show how the forecast changes over the hours of the day. We also realized that there was not an intuitive way to display both amount and portability of precipitation in a single QDP at a time. We changed our design to include 2 QDPs per hour - one for the amount of precipitation (inches per hour) and one for POP. Another aspect of the QDP that changed as we were implementing them was how we created the distribution. When we started generating our weather data based on real weather data, we realized that there would not be a way to have a real distribution of amount or probability of precipitation, so we decided to generate the distributions artificially.

Later in the design process, we also decided to coordinate the color coding for all three vizzes. We colored the amounts of rain categorically, so when each plot was showing heavy precipitation it would be dark blue, moderate would be medium blue, and light would be light blue.

All of these plots are described in more detail in the Implementation section.

7 Implementation

7.1 Text

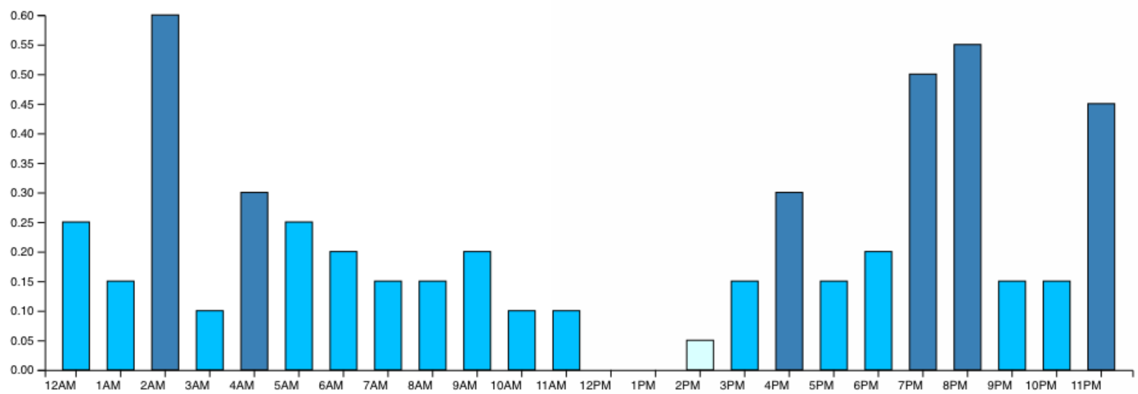
We used text describing the probability and amount of rain as a baseline against which to compare the performance of each visualization. The text described the probability of precipitation as well as the categorical value for the amount of precipitation (light, moderate, or heavy). Here is a sample of the text:

The following describes the chance of precipitation for tomorrow:

- At 12:00 am, there is a 25% chance of moderate precipitation.
- At 1:00 am, there is a 5% chance of light precipitation.
- At 2:00 am, there is no chance of precipitation.
- At 3:00 am, there is no chance of precipitation.
- At 4:00 am, there is no chance of precipitation.
- At 5:00 am, there is no chance of precipitation.
- At 6:00 am, there is no chance of precipitation.
- At 7:00 am, there is no chance of precipitation.
- At 8:00 am, there is a 10% chance of moderate precipitation.
- At 9:00 am, there is a 30% chance of moderate precipitation.
- At 10:00 am, there is a 45% chance of heavy precipitation.
- At 11:00 am, there is a 35% chance of moderate precipitation.
- At 12:00 pm, there is a 35% chance of light precipitation.
- At 1:00 pm, there is a 40% chance of moderate precipitation.
- At 2:00 pm, there is a 25% chance of light precipitation.
- At 3:00 pm, there is a 50% chance of moderate precipitation.
- At 4:00 pm, there is a 65% chance of moderate precipitation.
- At 5:00 pm, there is a 65% chance of moderate precipitation.
- At 6:00 pm, there is a 80% chance of heavy precipitation.
- At 7:00 pm, there is a 85% chance of moderate precipitation.
- At 8:00 pm, there is a 100% chance of moderate precipitation.
- At 9:00 pm, there is a 100% chance of heavy precipitation.
- At 10:00 pm, there is a 85% chance of heavy precipitation.
- At 11:00 pm, there is a 100% chance of moderate precipitation.

7.2 Static Bar Chart

We also used static bar charts to visualize the predicted rate of precipitation each hour the next day. These were built using D3.js and the data used was generated based on the methodology described in section 4. Each tick in the x-axis represented a different hour, and the bar was colored by the categorical level of precipitation (light, moderate, or heavy). These bar charts did not reference the chance of precipitation at all. We considered putting error bars on the bar graphs to visualize the chance of precipitation, but the chance of precipitation is not related to the rate of precipitation so including those error bars would not make sense. However, we still wanted to understand whether people generally preferred simple and easily readable vizzes like this bar chart that didn't show the uncertainty of the weather, or detailed vizzes that showed the uncertainty. An example of one of our static bar charts is shown below.



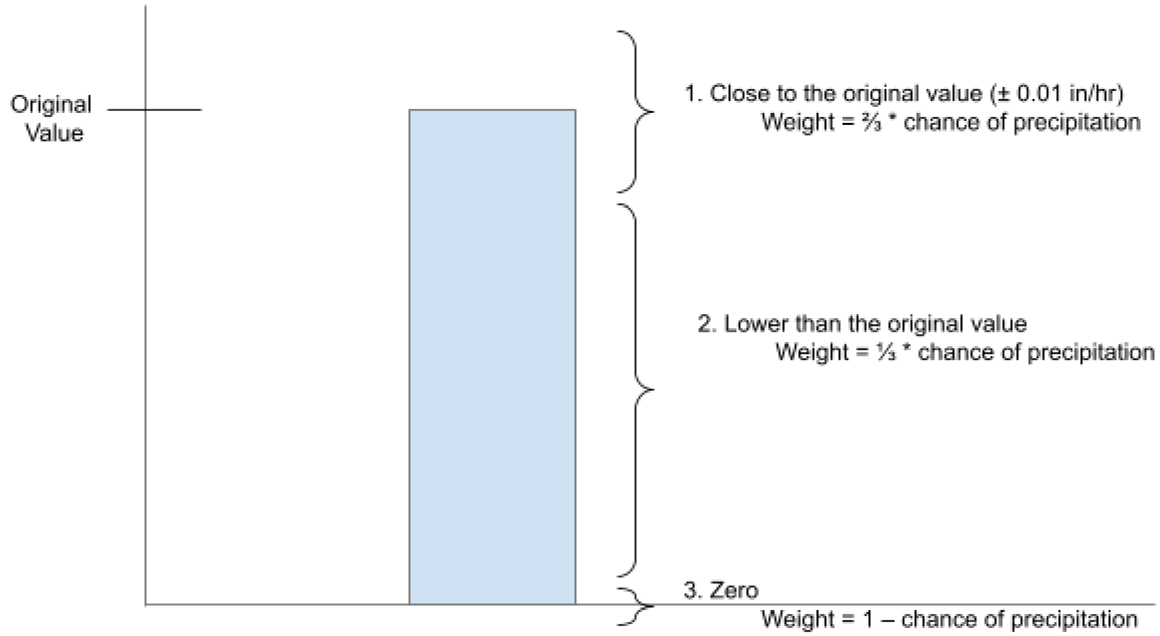
7.3 Hypothetical Outcome Plot

Another type of visualization we used were hypothetical outcome plots (HOPs), which are used to represent uncertain distribution dynamically rather than statically. The idea is to have a chart that periodically updates to show a new hypothetical draw from an uncertainty distribution, and the user can view this animation to mentally aggregate these draws and get a sense of the overall distribution. This is better than having bar graphs with error bars because when reading those charts, observers will typically make a decision based on the height of the bar graphs and do not account for the uncertainty that is shown by the error bars. HOPs, on the other hand, allow people to visually see how uncertain the data is, and observers who are trained to read them will have a better understanding of the uncertainty of the data than they would by reading bar graphs with error bars.

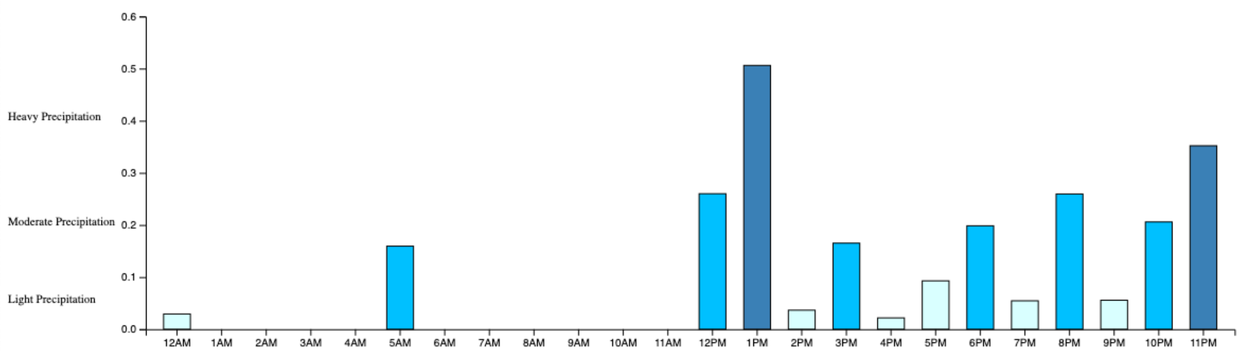
We built the HOPs using D3.js and essentially took the static bar chart animation and animated it every second to show a hypothetical prediction of what the day's weather might look like. We generated each set of new hypothetical predictions based on the chance of precipitation. For each hour, we performed a weighted selection on whether the new hypothetical prediction for that hour would be in one of three categories: close to the original prediction, lower than the original prediction, or zero. The weights for each hour were determined by the chance of precipitation:

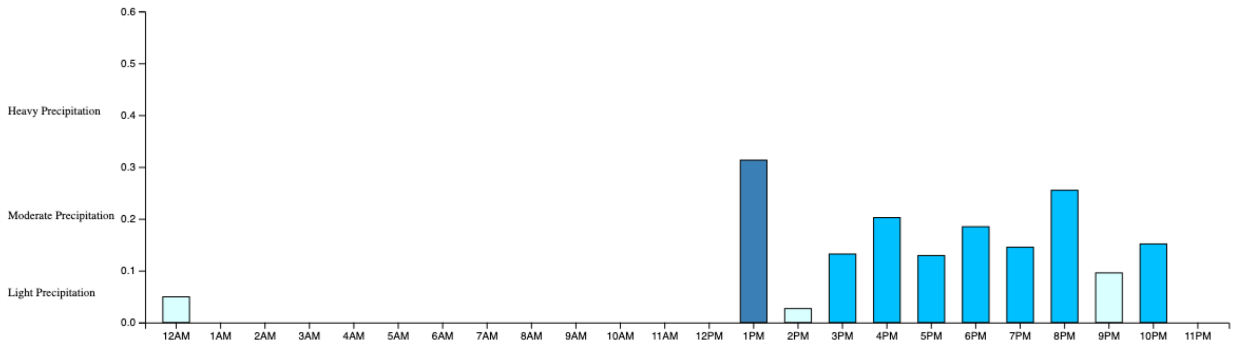
- Close to the original prediction: weight = ⅔ * chance of precipitation
- Lower than the original prediction: weight = ⅓ * chance of precipitation
- Zero: weight = 1 – chance of precipitation

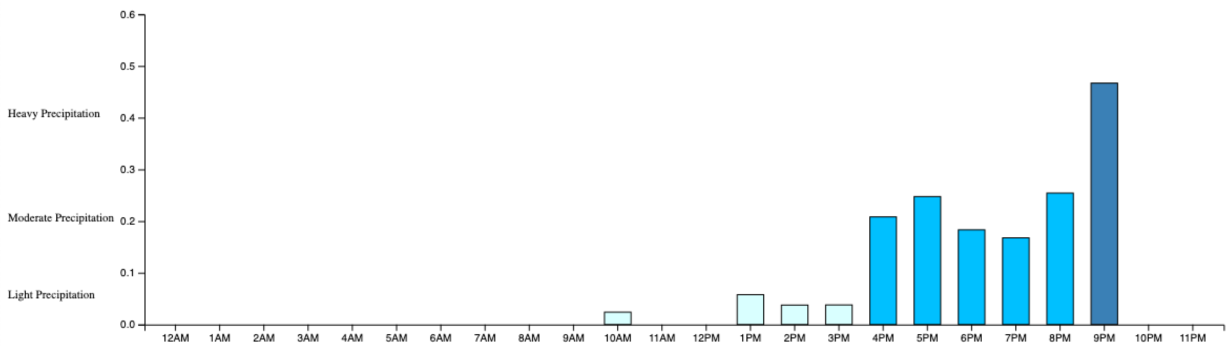
These weights were decided on because we wanted the probability of the new hypothetical data being zero to be similar to the uncertainty of precipitation. Likewise, we wanted the probability of the data to fall somewhere near the predicted value to be more than the probability of it being another value. These weights are also summarized in the following figure.



Here are three different iterations for the same HOP. Like the static bar charts, they are colored based on the level of precipitation the value is under (light, moderate, heavy).

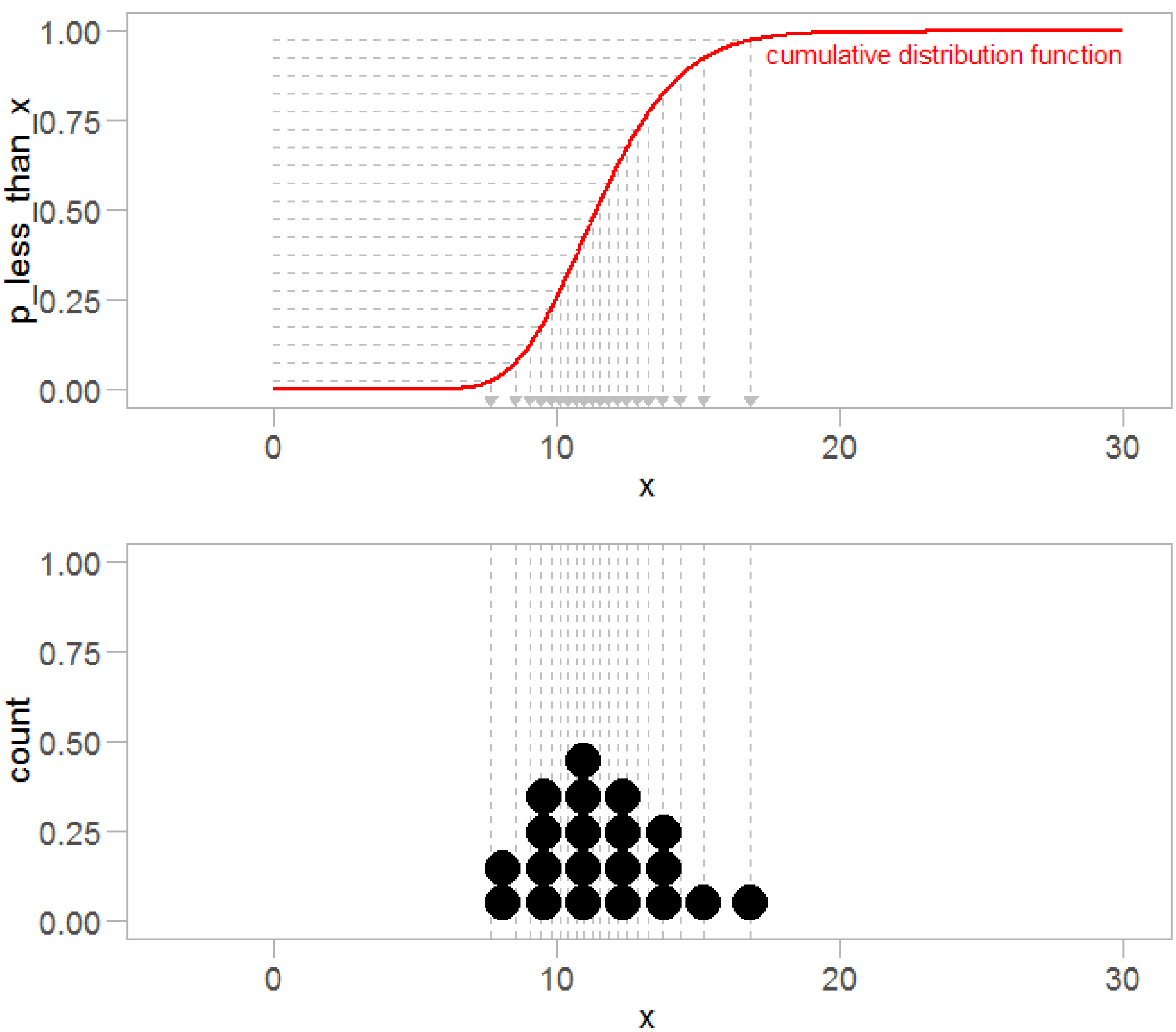




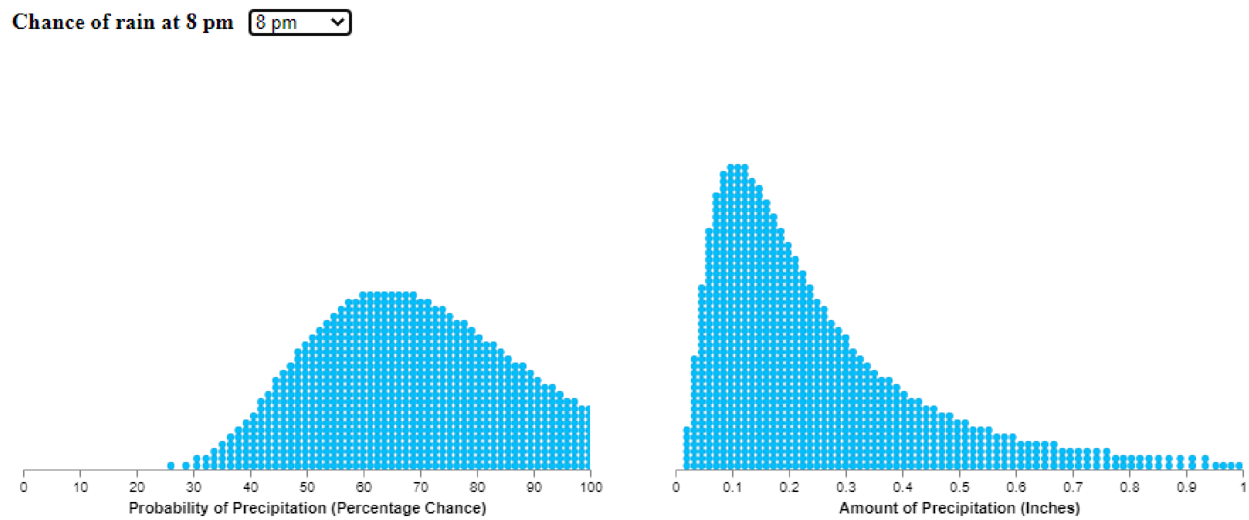


7.4 Quantile Dot Plot

Our final visualization was a *quantile dot plot*. A quantile dot plot is typically used as a way to represent continuous probability distributions in a discrete form. Kay et Al. in their paper "When(ish) is My Bus?" propose this vis idiom as a representation of uncertainty due because discrete outcomes are often easier for humans to interpret. To create a quantile dot plot, first the cumulative distribution function, is converted into a number of quantiles. Then the quantiles are drawn as a dot plot. As an example from [Matt Kay's github](), you can see how a CDF can be translated into a dot plot.



These plots typically represent uncertainty data for one dimension at a specific time. For example, as the paper's title implies, "When(ish) is My Bus?" uses quantile dot plots to represent the distribution of possible actual arrival times for a bus around the scheduled time. In our experiment we are representing more dimensions of uncertainty. For a given hour in a day, we want to represent the chance of precipitation as well as the amount of precipitation. To accommodate multiple dimensions, we show two side by side quantile dot plots, as seen in the figure below. The dot plot on the right shows the distribution of the probability of precipitation. The dot plot on the right shows the distribution of possible amounts of precipitation. More dots on a value mean that value is more likely. So to interpret the figure below one might say there is about a 65% chance of getting about 0.15 inches of rain at 8pm. We allowed users to choose the time they wanted to view distributions for, or let the graphs animate between the different hours. The animations have the option of very fast, 1, 2, or 5 second intervals.



To implement this visualization we used Vega, inspired by [this tutorial](). Our data, as described in Section 4, generates hourly forecast data with a probability and an amount of precipitation. Vega includes a function to generate quantiles for a normal distribution from an average and standard deviation value. Using our generated data as the average, we did not have our own CDF, so we set a standard deviation that created interesting, interpretable plots and used this Vega function to create our dot plots. The Vega tutorial we used as our initial guide was based on the "When(ish) is My Bus?" paper so it seemed like a useful approach for generating our distributions. We used the Vega-embed library to embed this visualization into our application.

7.5 React and Firebase

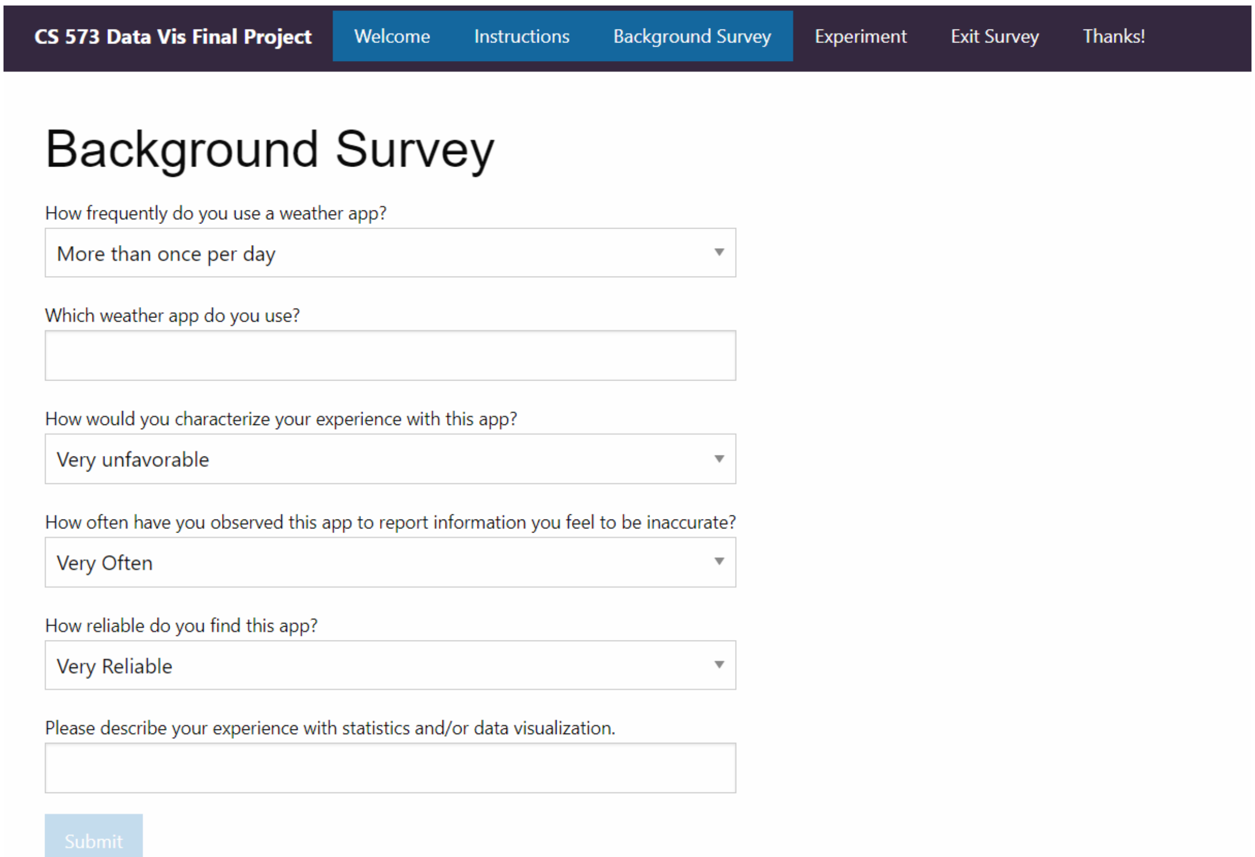
We designed our experiment to be a React web application. Due to the limited time available for development during the final project, this allowed us to reuse code created for our survey in A3. React also allowed for our website to be a modern, responsive web-app. Reusable React components helped our code stay clean and allowed us to provide many random trials to the user in a smooth, controlled way. The Foundation styling library was used to give the web app a stylish, modern feel that stayed consistent on all pages and devices.

We needed a way to make our experiment available to the public. Firebase provides free and easy hosting for React web applications. We leveraged these features to create https://cs573-finalproject.web.app/. This website uses Firebase hosting to host the site at a custom URL based on our project name. Since we were already using Firebase for hosting, we also used their *Realtime Database* feature to store our data. Firebase Realtime Database provides a NoSQL storage solution for data. This works very nicely with JSON data and React. In our React app as users take the survey, their results are stored in the React state for the application. Using the Firebase API, once the survey is complete we upload the React state (a JSON object) to the Realtime Database, which then converts the JSON into NoSQL to store. After the survey was complete, Firebase provides an option to export the database contents as a JSON file which we could then examine using other tools.

8 User Study Design

To determine the effectiveness of the different visualizations for displaying uncertainty information in weather data, we developed a survey to measure the relative effectiveness and cognitive workload of the different visualizations described in Section 7. We also collected information regarding the background of the users taking the survey. The survey begins by asking users six background questions:

- How frequently do you use a weather app?
- Which weather app do you use?
- How would you characterize your experience with this app?
- How often have you observed this app to report information you feel to be inaccurate?
- How reliable do you find this app?
- Please describe your experience with statistics and/or data visualization.



Next, we provided users with a series of 20 questions, (5 per vis type). Each of these five questions will be different, two are visualization readability questions, and three are decision making questions. Due to time constraints and encouraging users to respond, we kept the survey short. As it is a pilot study, we thought these questions would provide an adequate experiment for a pilot study of weather uncertainty visualization and we can use the results of this data to determine future directions for uncertainty research.

Here are the five different question types we present for each visualization:

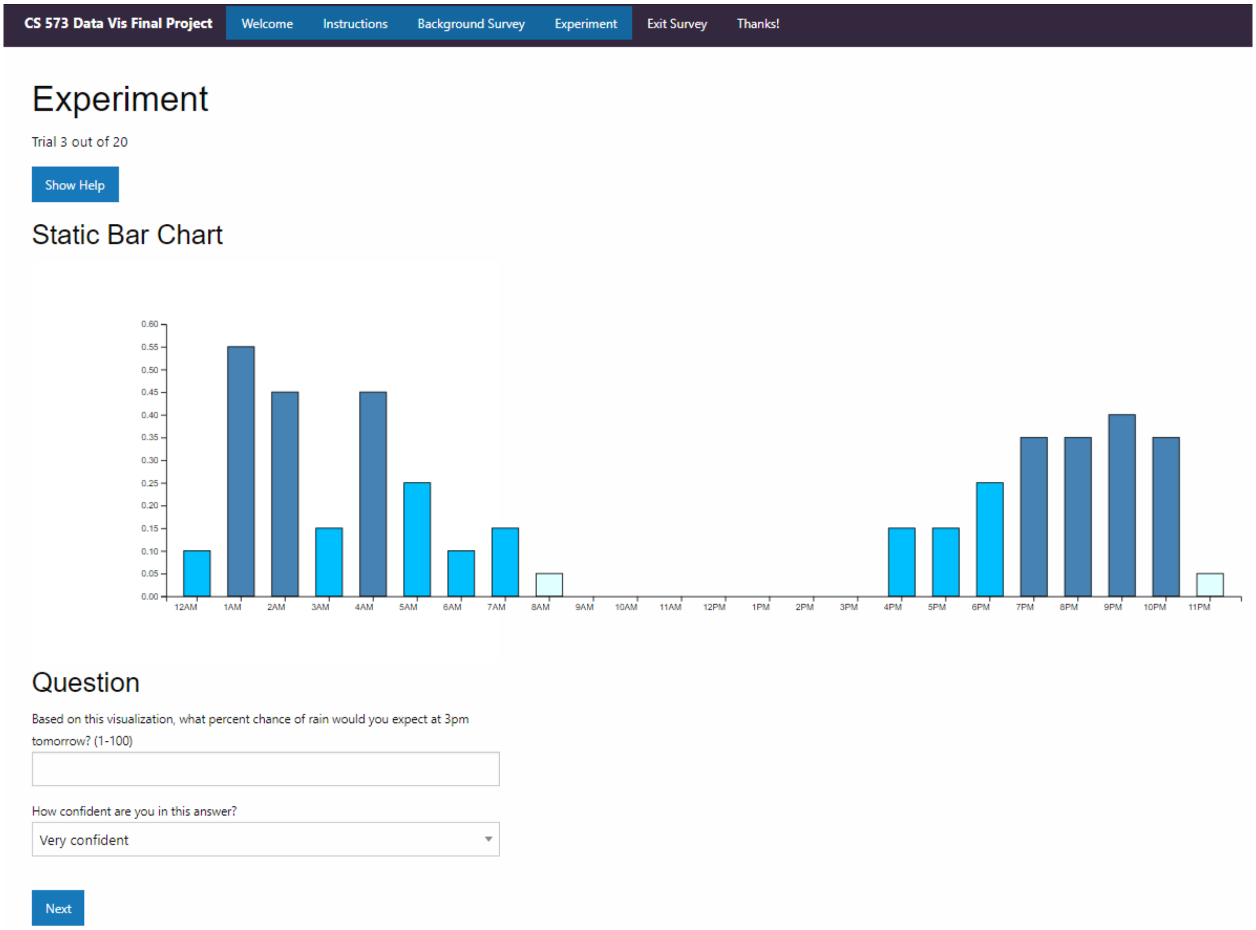
Decision-Making Questions

- If you had scheduled a hike with a friend for 3pm tomorrow, and the following forecast for tomorrow was presented to you, would you reschedule?
    - How confident are you in your decision?
    - Please briefly describe the process you used to make your decision, including how you interpreted the visualization.
- Assuming that you enjoy taking afternoon walks and would have availability, would you plan to go outside at 3pm tomorrow, based on this information?
    - How confident are you in your decision?
    - Please briefly describe the process you used to make your decision, including how you interpreted the visualization.
- You are in charge of planning an extracurricular club event or a community event that is currently to be held outdoors. This event has been planned for a few weeks, and you have neglected to book an indoor location as a backup. The day before the event, you saw the following forecast for the day of the event. Would you pay a small fee to book the late space at the last minute in case it rains?
    - How confident are you in your decision?
    - Please briefly describe the process you used to make your decision, including how you interpreted the visualization.

Visualization Readability Questions

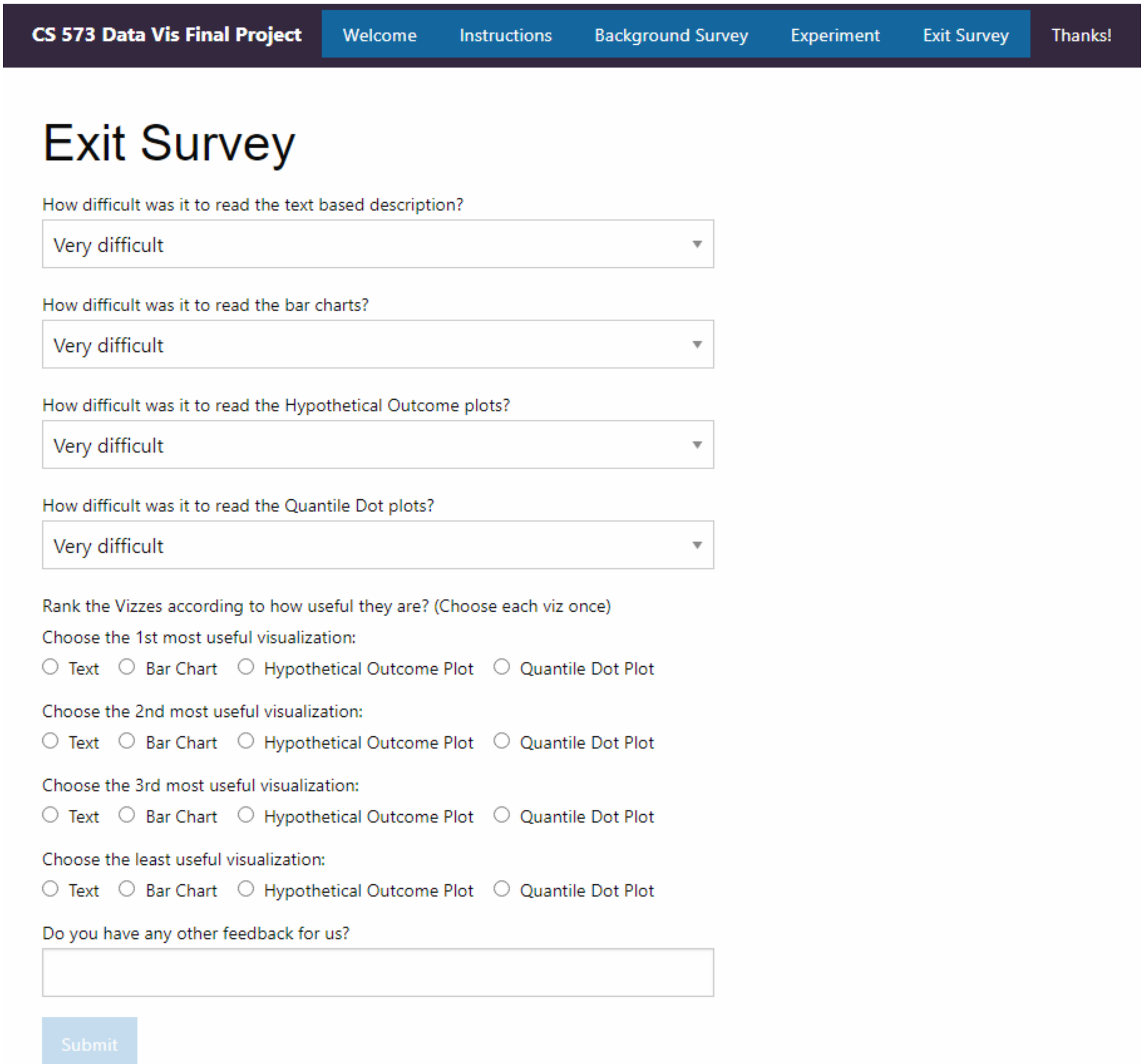
- Based on this visualization, how much rain do you expect at 3pm tomorrow? (Answer in inches per hour)
    - How confident are you in this answer?
- Based on this visualization, what percent chance of rain would you expect at 3pm tomorrow?
    - How confident are you in this answer?

The questions and visualizations are presented in a random order and the weather data is generated at the start of each of the twenty trials. The following is a screenshot of one trial in the experiment showing a bar chart visualization with a readability question asking about the probability of precipitation. Each experiment also includes a “Show Help” button. We track if users ask for help on each trial.



More examples of our survey questions can be seen in our survey, https://cs573-finalproject.web.app/, or in Appendix 3: Example Survey Questions.

After completing the twenty trials the users are presented with an exit survey. This survey asks them two very important questions, the difficulty required to interpret each visualization and a ranking of the relative usefulness of each visualization. We also provide the users with a textbox to provide any additional feedback about the visualizations and weather uncertainty.



After the exit survey is submitted, the response will be stored in a Firebase Realtime Database and the user is prompted to take the survey again. Taking the survey a second time will only present the users with the experiment and exit survey, the background questions do not need to be filled out more than once. The structure of the survey data we collected can be seen in Appendix 2: Example Survey Response JSON.

9 Exploratory Data Analysis

Once we had the results from the experiment, we did some data analysis on the data we collected. We visualized and quantitatively analyzed the data we collected in order to determine trends in the data and compare the different vis types. We used primarily Tableau to produce these visualizations, as well as Google Sheets for two stacked bar charts. The Tableau file can be viewed on Tableau Public at this link.

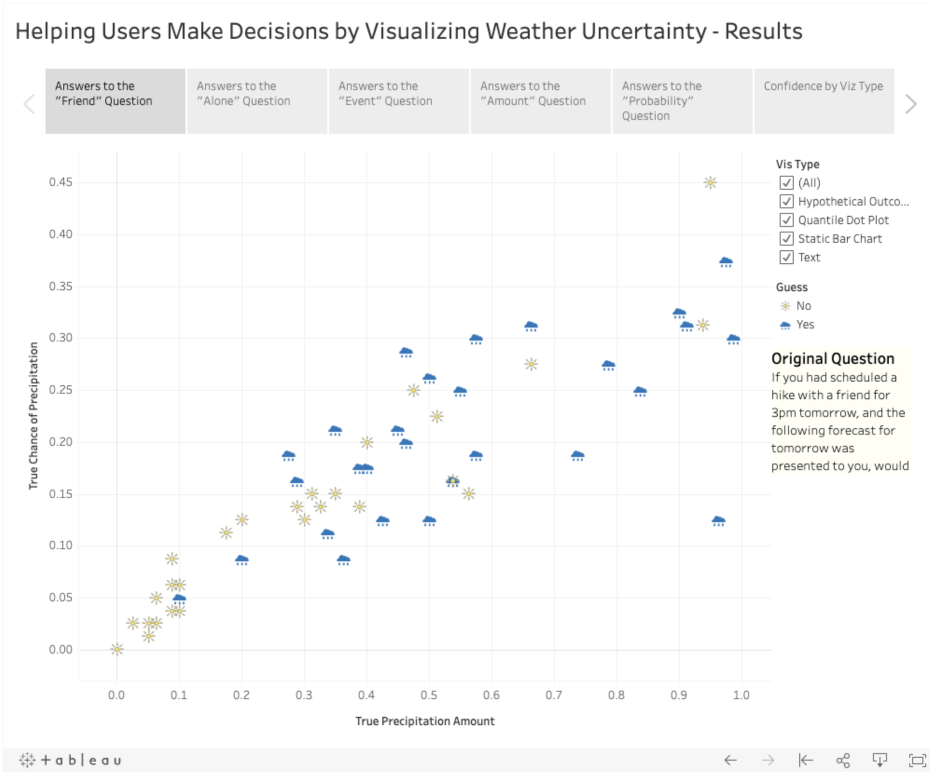
9.1 Trial Questions Results

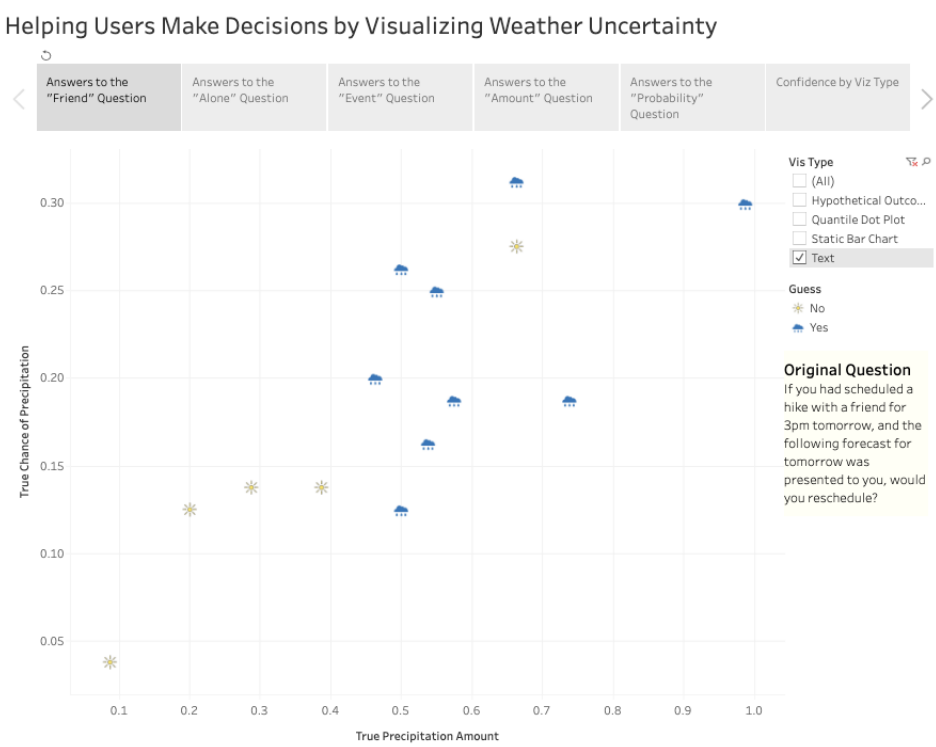
For each of the 5 question types, we visualized the answers that the users gave. For the "friend" and "alone" questions, we plotted the weighted average amount and weighted probability of precipitation over the 5 hours around 3pm (1pm-5pm), with 3pm more heavily weighted, as the "true amount of precipitation" and "true chance of precipitation" on the x and y axes, respectively. We assumed that the participant might have been looking at the weather conditions before the hike to determine, for example, whether the trail would be muddy or slippery. In addition, we assumed that the participant would also look at the weather conditions after 3pm because hikes typically last for a couple hours or so. Thus, when calculating the weighted averages, 3pm was given a 20% weight, 3pm was given a 50% weight, and 3-5pm was given a 30% weight. For all three questions with yes or no answers, we encoded the answer to the question ("Yes" or "No") in the scatterplot as the marker types:

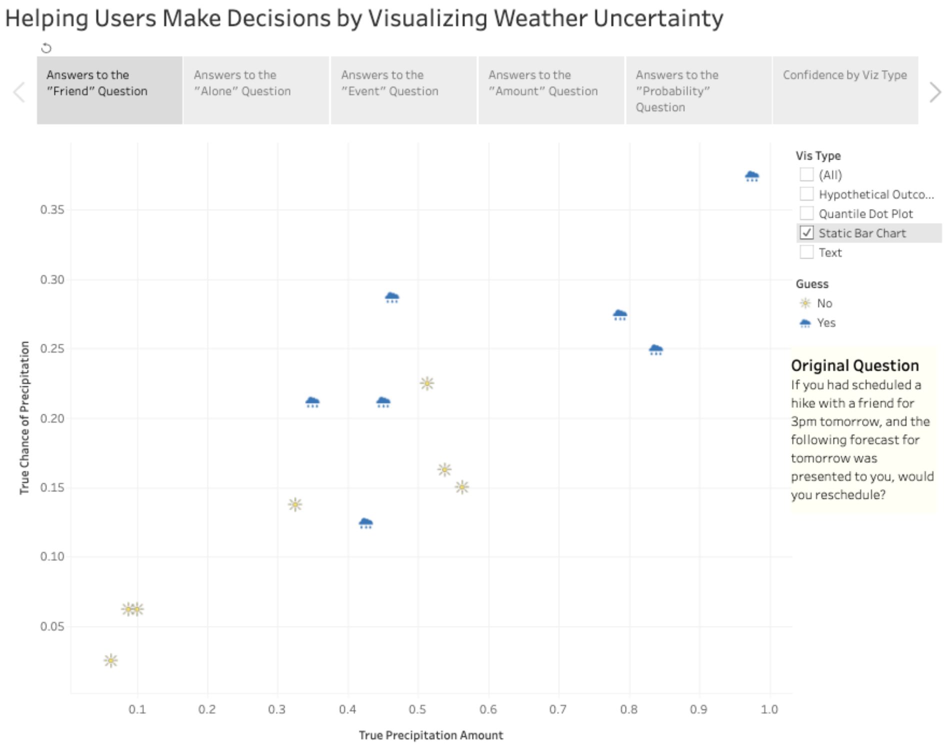
- For the "friend" question:
    - "No" (I would not reschedule) is a sun
    - "Yes" (I would reschedule) is a rain cloud
- For the "alone" question:
    - "No" (I would not go for a walk) is a rain cloud
    - "Yes" (I would go for a walk) is a sun
- For the "event" question:
    - "No" (I would not book an indoor venue) is a sun
    - "Yes" (I would book an indoor venue) is a rain cloud

These plots are available on Tableau Public, where you can also filter by the vis type. We have included screenshots of them here:

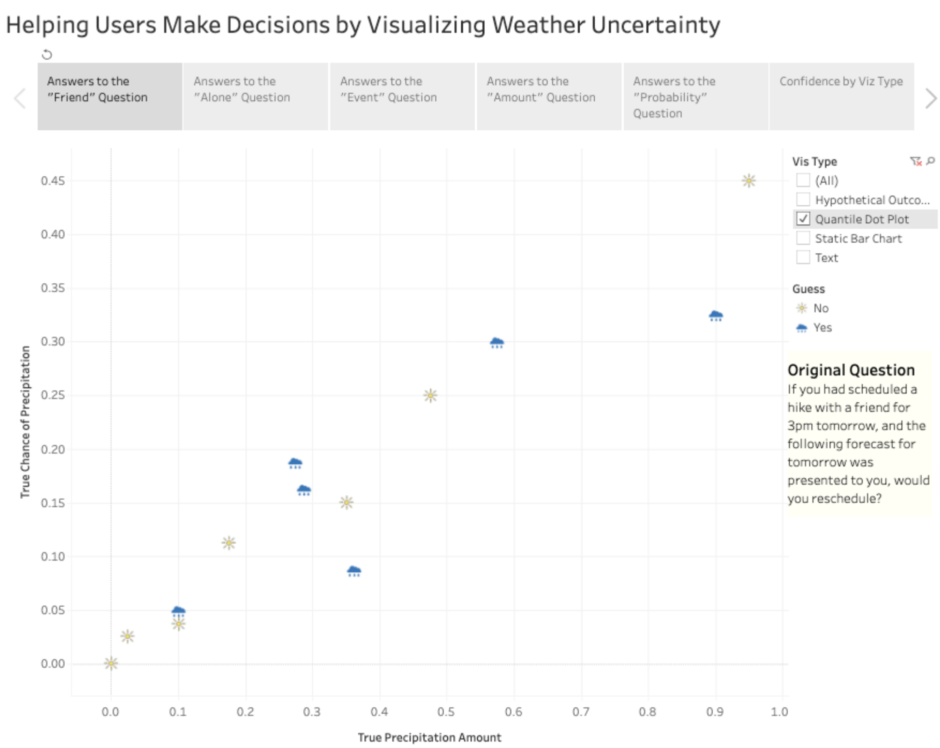
This following plot graphs the "true precipitation amount" against the "true chance of precipitation" for the "friend" question. For this question, we have also included screenshots of how the graph looks when it is filtered to include each of the viz types. The ideal results should have a cluster of yellow suns near the bottom left of the graph – if there is a low chance of precipitation and a low predicted rate of rainfall, participants should ideally answer "No" to rescheduling – and a cluster of blue rain clouds near the top right of the graph – if there is a high chance of precipitation and a high predicted rate of rainfall, participants should ideally answer "Yes". Based on these figures, text had the clearest and most defined clusters, followed by bar plots, HOPs, and finally quantile dot plots.


Helping Users Make Decisions by Visualizing Weather Uncertainty - Results
Answers to the "Friend" Question
Answers to the "Alone" Question
Answers to the "Event" Question
Answers to the "Amount" Question
Answers to the "Probability" Question
Confidence by Viz Type
True Chance of Precipitation
True Precipitation Amount
Vis Type
(All)
Hypothetical Outco...
Quantile Dot Plot
Static Bar Chart
Text
Guess
No
Yes
Original Question
If you had scheduled a hike with a friend for 3pm tomorrow, and the following forecast for tomorrow was presented to you, would
tableau

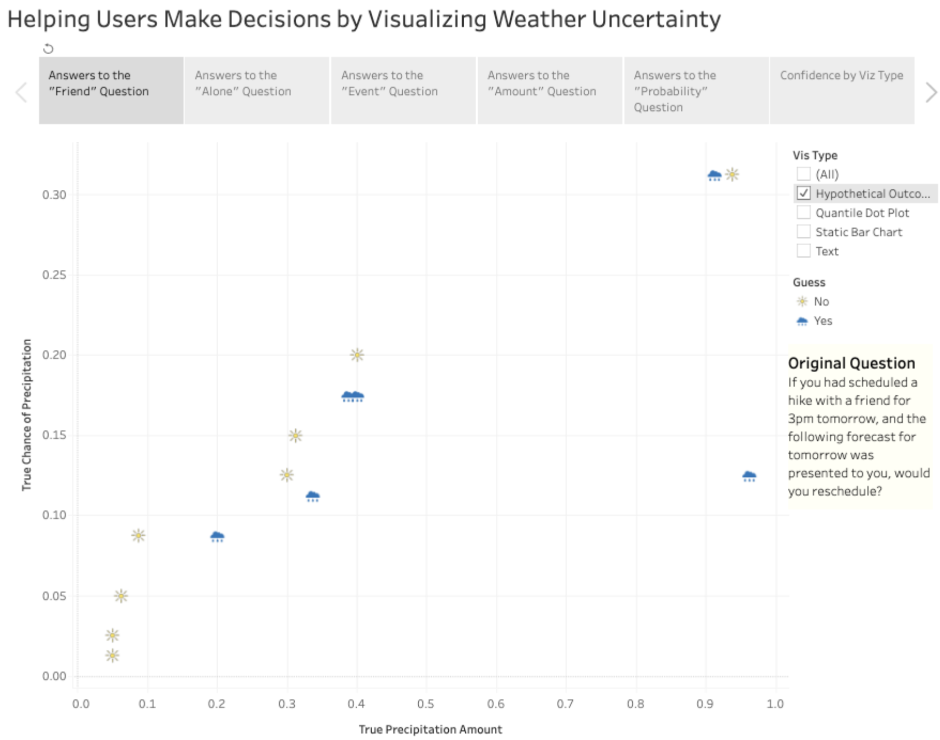

Helping Users Make Decisions by Visualizing Weather Uncertainty
Answers to the "Friend" Question
Answers to the "Alone" Question
Answers to the "Event" Question
Answers to the "Amount" Question
Answers to the "Probability" Question
Confidence by Viz Type
True Chance of Precipitation
True Precipitation Amount
Vis Type
(All)
Hypothetical Outco...
Quantile Dot Plot
Static Bar Chart
Text
Guess
No
Yes
Original Question
If you had scheduled a hike with a friend for 3pm tomorrow, and the following forecast for tomorrow was presented to you, would you reschedule?



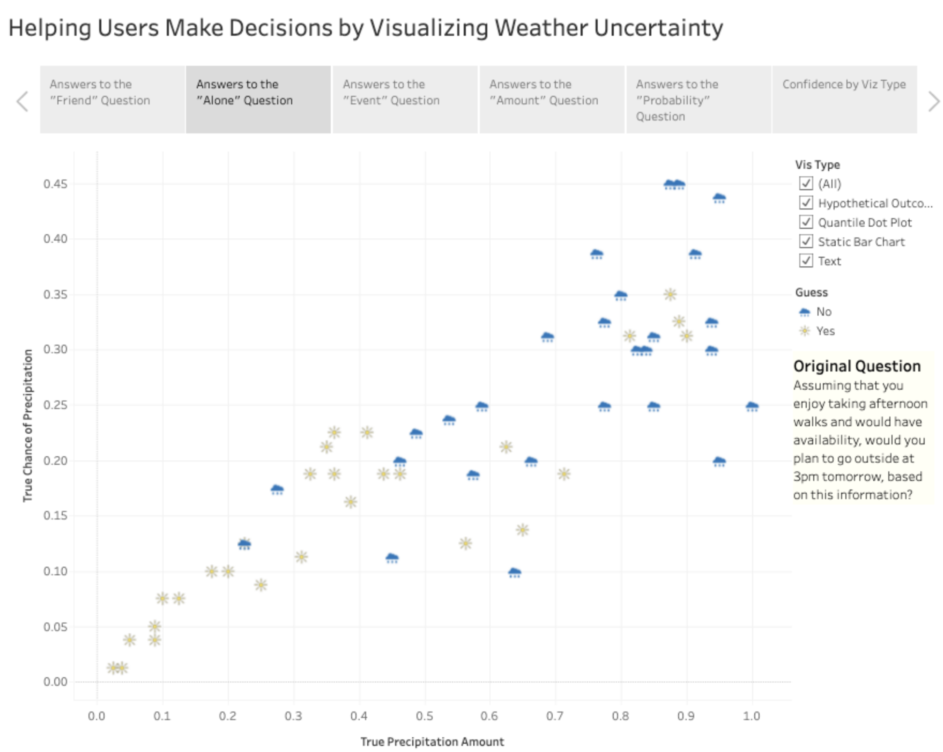
Helping Users Make Decisions by Visualizing Weather Uncertainty
Answers to the "Friend" Question
Answers to the "Alone" Question
Answers to the "Event" Question
Answers to the "Amount" Question
Answers to the "Probability" Question
Confidence by Viz Type
Vis Type
(All)
Hypothetical Outco...
Quantile Dot Plot
Static Bar Chart
Text
Guess
No
Yes
Original Question
If you had scheduled a hike with a friend for 3pm tomorrow, and the following forecast for tomorrow was presented to you, would you reschedule?
True Chance of Precipitation
True Precipitation Amount



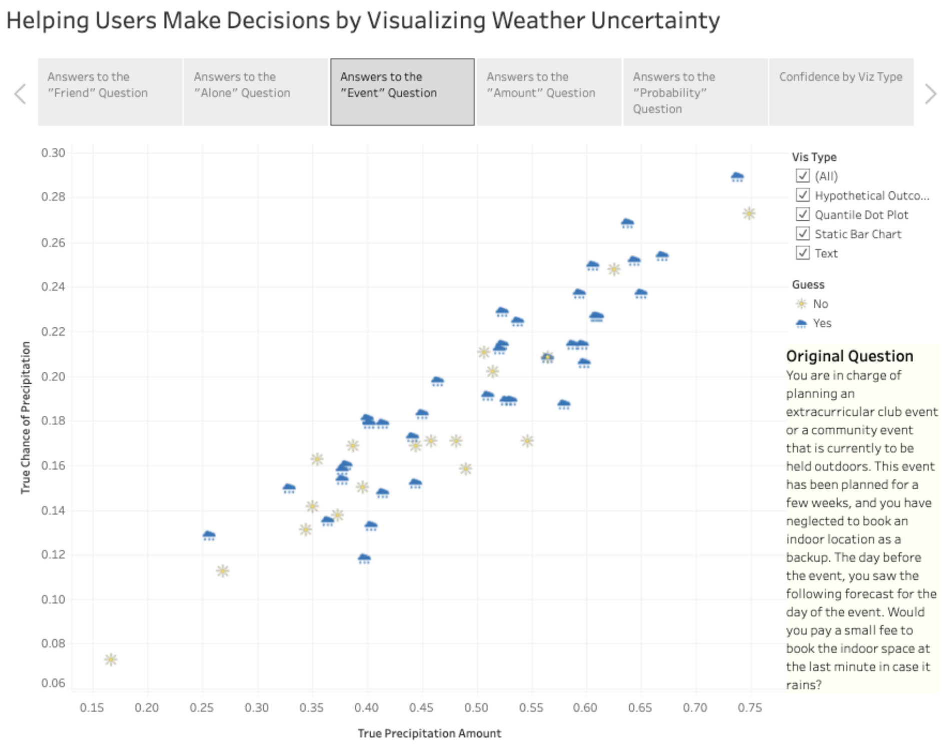
Helping Users Make Decisions by Visualizing Weather Uncertainty
Answers to the "Friend" Question
Answers to the "Alone" Question
Answers to the "Event" Question
Answers to the "Amount" Question
Answers to the "Probability" Question
Confidence by Viz Type
Vis Type
(All)
Hypothetical Outco...
Quantile Dot Plot
Static Bar Chart
Text
Guess
No
Yes
Original Question
If you had scheduled a hike with a friend for 3pm tomorrow, and the following forecast for tomorrow was presented to you, would you reschedule?
True Chance of Precipitation
True Precipitation Amount



The following figure shows the results for the “alone” question. For this question, the most defined clusters were from the text, followed by the quantile dot plots, and the HOPs and static bar charts had similar unclear clusters.

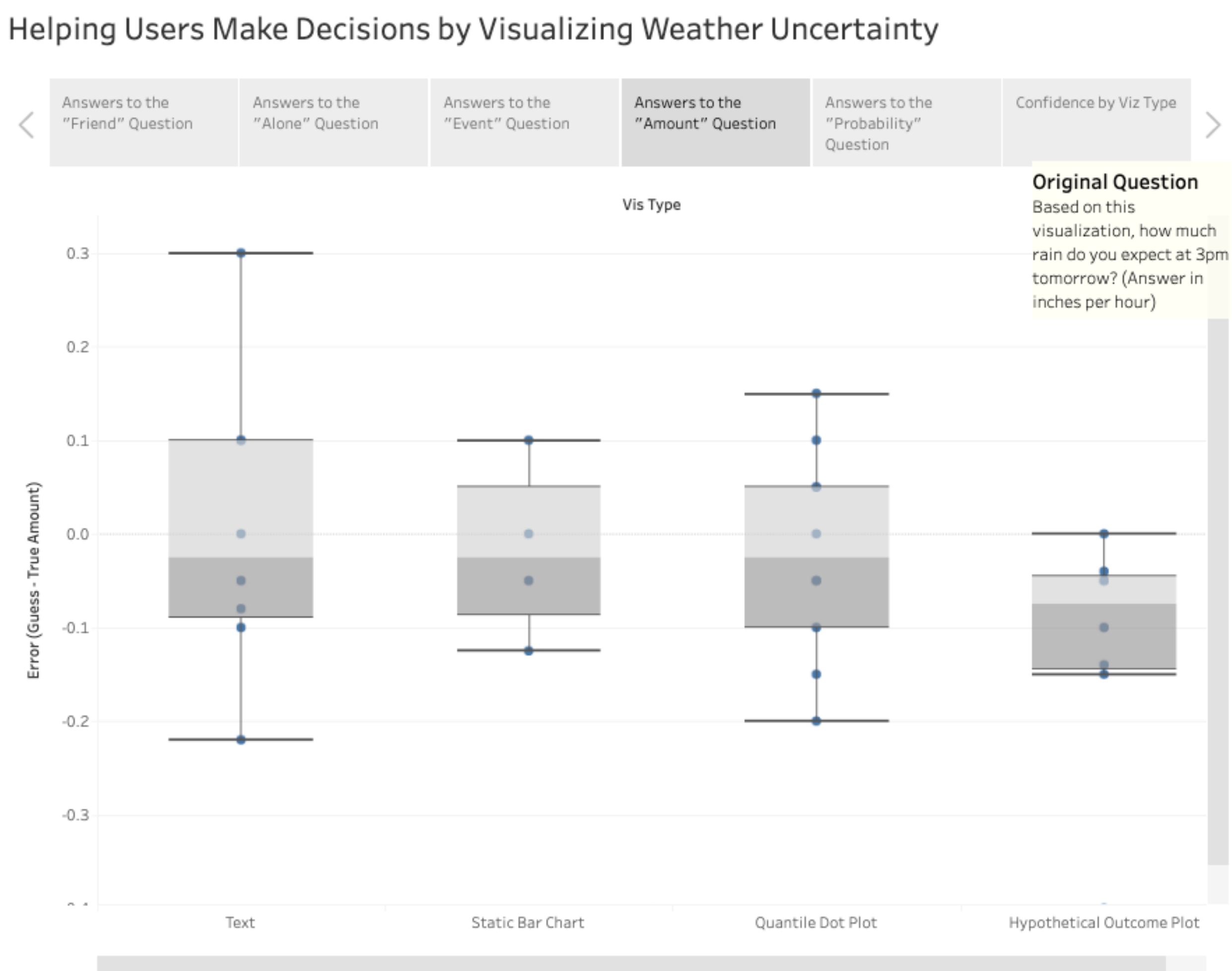


For the “event” question, we used the average amount and probability of precipitation over the entire day as the “true amount of precipitation” and “true chance of precipitation” values on the x and y axes, respectively, since this question references an all-day event. This plot is also available on [Tableau Public](), where you can also filter by the vis type. We have included a screenshot of it here:

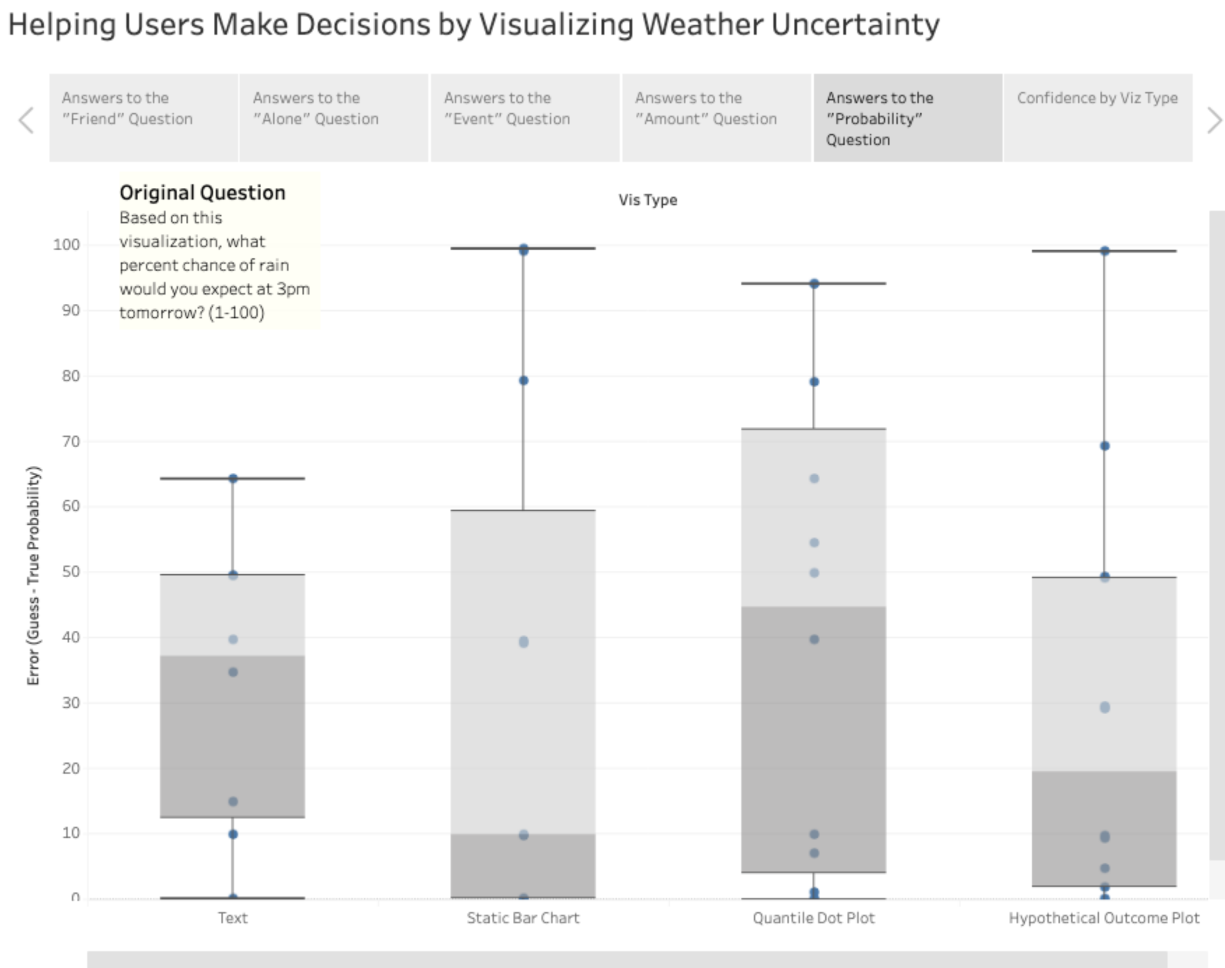


Almost everyone answered “Yes I would book an indoor venue” to the event question when given the text or quantile dot plots. There was some clustering in the static bar chart, and some slight divisions in the hypothetical outcome plots.

There were also questions that asked for quantitative responses about the probability and amount of rain expected. We computed the error between the guess and the true amount or probability of rain at the hour in question, and plotted these errors as box plots by vis type.



When trying to predict the amount of rain, the average guesses were consistently below the actual amount of rain at that hour, regardless of vis type. However, there was a much wider range (more variability) in the responses when looking at the text. This could be because the text only specified “light”, “moderate”, and “heavy” precipitation whereas the other plots showed numeric measures as well as the categories.

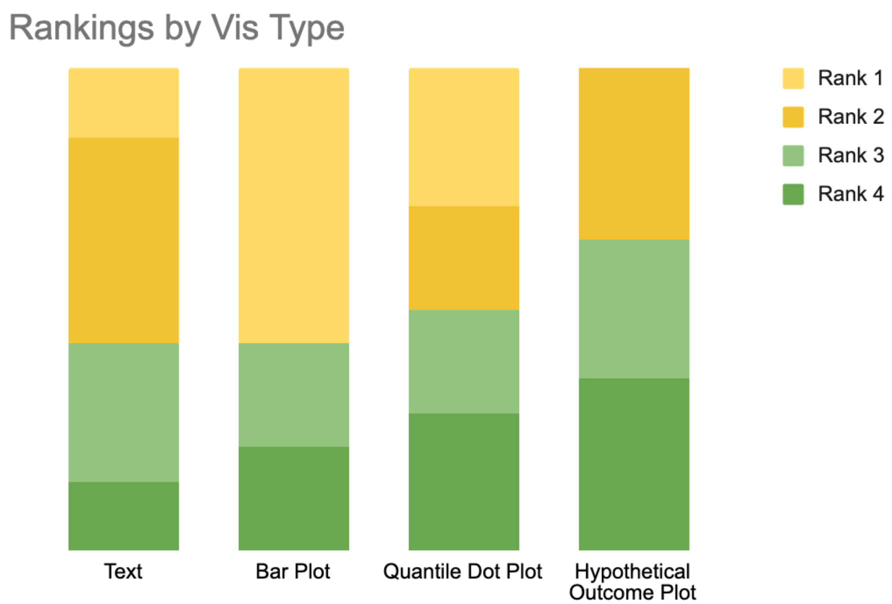


When guessing the probability of rain at a given time, users had significant error compared to the actual POP values. For static bar charts and hypothetical outcome plots, the error was even up to 100%. The errors were lowest for the static bar charts than for the text and second lowest for the HOPs.

9.2 Exit Survey Results

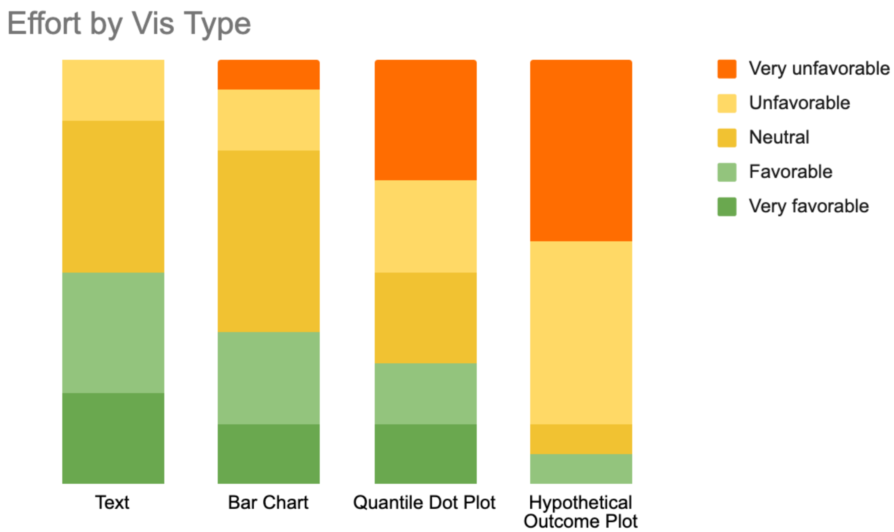
At the end of our experiment, we asked participants several questions to summarize their experience with the vizzes.

We asked participants to rank which vizzes were their favorites. The results from this question are visualized in the figure below:



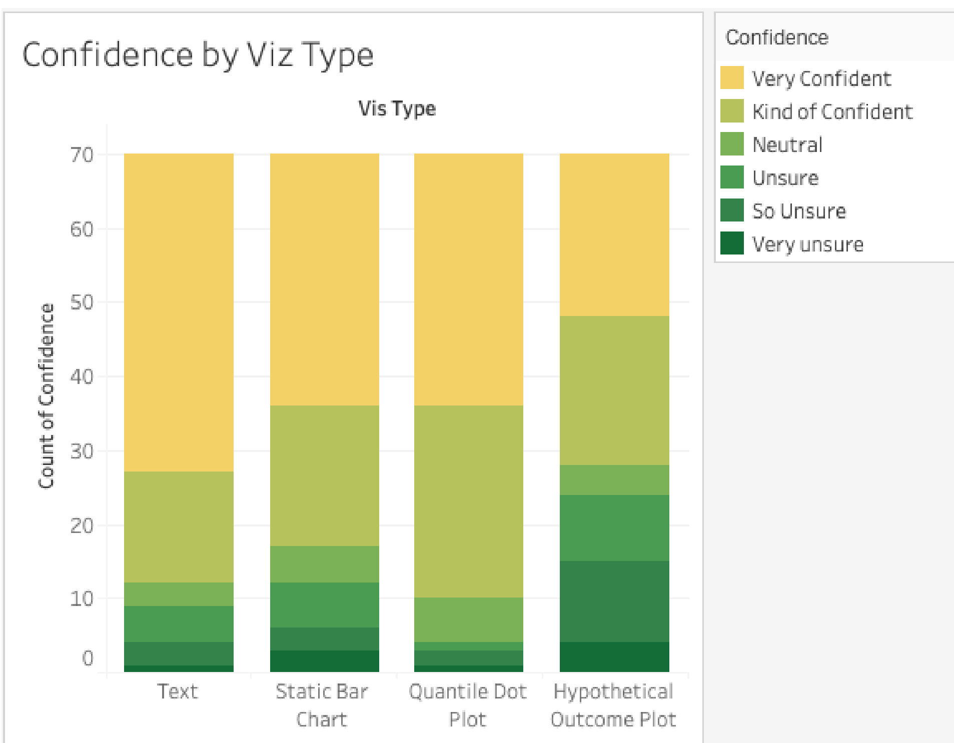
This stacked bar chart shows that participants were most likely to rank the bar plot highest. The text also had about 50% of participants ranking it in the upper half. The hypothetical outcome plots were ranked the worst - they were nobody's favorite visualization, and they had the largest number of rankings at the bottom (Rank 4).

We also asked participants to describe how much effort they had to put into reading each vis type. "Effort" is a subjective metric that is up to the participant to describe. The results of this questions are visualized below:



This stacked bar chart shows that participants felt that the text required the least effort to read, followed by the static bar charts. The hypothetical outcome plots took the most effort for participants to read.

We also asked participants to describe how confident they were with each answer they provided, including answers to all 5 question types. The results from these questions are visualized in the following figure:



These results show that participants were slightly more confident in their answers when looking at text, followed by quantile dot plots. They were least confident in their answers when looking at hypothetical outcome plots.

10 Evaluation

10.1 Results of User Study

In our experiment, we asked the participants five different questions that fit into two categories: decision-based questions, and visualization readability questions. The visualization readability questions – the "amount" question and the "probability" question asked the participant to guess either the amount of precipitation or the chance of precipitation that the graph was showing.

Exit Survey: Most Preferred Vis Types				
	Best/Highest (1)	**Second best (2)**	**Second worst (3)**	**Worst (4)**
Effort	Text	Static bar chart	Quantile dot plot	Hypothetical outcome plot
Ranking	Static bar chart	Text	Quantile dot plot	Hypothetical outcome plot
Confidence	Text	Quantile dot plot	Static bar chart	Hypothetical outcome plot

When looking at these results, we have to take into account that the users liking a vis does not mean that that vis does the best job at accurately conveying the information. In the following table we look more closely at how accurately the participants perceived the information for each vis type:

Trial Questions: Accuracy by Vis Type				
	Best/Highest (1)	**Second best (2)**	**Second worst (3)**	**Worst (4)**
Clusterability: Friend Question	Text	Static bar chart	Hypothetical outcome plot	Quantile dot plot
Clusterability: Alone Question	Text	Quantile dot plot	Static bar chart / Hypothetical outcome plot	Static bar chart / Hypothetical outcome plot
Clusterability: Event Question	Static bar chart	Hypothetical outcome plot	Quantile dot plot	Text

Accuracy: Amount Question	Static bar chart	Quantile dot plot	Text	Hypothetical outcome plot
Accuracy: Probability Question	Static bar chart	Hypothetical outcome plot	Text	Quantile dot plot

As mentioned above, text may have performed poorly on the quantitative questions because it only provided the category of precipitation ("light", etc) whereas the other vizzes provided some numeric information about the POPs and amounts of precipitation.

To compare the vizzes against each other, we computed their "scores" by taking their average score (1-4) across the rows of each table above. Higher scores indicate worse performance or ratings.

Vis Type	Score on User Preference	Score on Accuracy
Text	1.33	2.4
Static Bar Chart	2	1.7
Quantile Dot Plot	2.66	3
Hypothetical Outcome Plot	4	2.9

This summary of our findings shows that the users preferred the text and static bar charts most and the HOPs least. The static bar charts performed the best overall for accuracy of user's perceptions, while the HOPs and QDPs were not as good for accuracy. Overall, a combination of text and static bar charts, as is used on the most popular weather apps from our survey, could result in the best outcomes for users.

11 Conclusions

In conclusion, our pilot study provided results about which weather uncertainty visualizations we developed were the easiest to read and did the best job at conveying the uncertainty of precipitation. We found that hypothetical outcome plots were difficult to read and unfavorable for participants. Quantile dot plots were slightly more favorable to participants, but slightly worse for accuracy of perceptions. Text was very easy to read and left participants feeling sure of their decisions and perceptions, but it had worse accuracy than static bar charts. Many participants said static bar charts were their favorite of the vizzes we presented, and bar charts scored the best on accuracy.

11.1 Future Work

This study was only a pilot study intended to get an idea about which vizzes might be best for visualizing weather uncertainty for decision making. In the future this study could be expanded to be more complete. This could involve using real precipitation data rather than artificially generated data, having more participants take the survey so statistically significant results can be drawn, and doing more rigorous calculations of the distributions for the quantile dot plots and hypothetical outcome plots. It could also be expanded to include more types of vizzes, such as some weather uncertainty vis that includes value suppressing uncertainty palettes. We also recommend for a future study to try out combinations of these vizzes. For example, it could test whether combining text with a static bar chart performs as well as the text and bar chart individually.

In addition, since the goal of this study was to determine the best weather uncertainty visualizations, another step that could be taken in the future would be to implement the results from this study and future weather uncertainty vis studies into a weather forecast app that is optimally designed to communicate precipitation uncertainty to users.

References

1. 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*, 2016, pp. 5092–5103.
2. A. Kale, F. Nguyen, M. Kay, and J. Hullman, “Hypothetical outcome plots help untrained observers judge trends in ambiguous data,” *IEEE transactions on visualization and computer graphics*, vol. 25, no. 1, pp. 892–902, 2018.
3. M. Correll, D. Moritz, and J. Heer, “Value-suppressing uncertainty palettes,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1–11.
4. J. Sanyal, S. Zhang, J. Dyer, A. Mercer, P. Amburn, and R. Moorhead, “Noodles: A tool for visualization of numerical weather model ensemble uncertainty,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 6, pp. 1421–1430, 2010.
5. L. Nadav-Greenberg, S. L. Joslyn, and M. U. Taing, “The effect of uncertainty visualizations on decision making in weather forecasting,” *Journal of Cognitive Engineering and Decision Making*, vol. 2, no. 1, pp. 24–47, 2008.
6. A. Kale, M. Kay, and J. Hullman, “Visual reasoning strategies and satisficing: How uncertainty visualization design impacts effect size judgments and decisions,” *arXiv preprint arXiv:2007.14516*, 2020.
7. J. Hullman, X. Qiao, M. Correll, A. Kale, and M. Kay, “In pursuit of error: A survey of uncertainty visualization evaluation,” IEEE transactions on visualization and computer graphics, vol. 25, no. 1, pp. 903–913, 2018.
8. J. Hullman, "Why Authors Don't Visualize Uncertainty," in IEEE Transactions on Visualization and Computer Graphics, vol. 26, no. 1, pp. 130-139, Jan. 2020, doi: 10.1109/TVCG.2019.2934287
9. Weather Shack, “Rain Measurement,” 2021, https://www.weathershack.com/static/ed-rain-measurement.html
10. OpenWeather API, 2021, https://openweathermap.org/api
11. J. Evans, “Creating a Production Build”, 2019, https://create-react-app.dev/docs/production-build/
12. AV Dojo, “React and Firebase | Firebase Realtime database with React |”, 2019, https://www.youtube.com/watch?v=0pC8dEqSKkc
13. J. Richards, “How to Deploy a React App with Firebase Hosting”, 2019, https://medium.com/swlh/how-to-deploy-a-react-app-with-firebase-hosting-98063c5bf425
14. Vega, “Quantile Dot Plot Example”, https://vega.github.io/vega/examples/quantile-dot-plot/
15. M. Kay, “Quantile dotplots”, 2016, https://github.com/mjskay/when-ish-is-my-bus/blob/master/quantile-dotplots.md

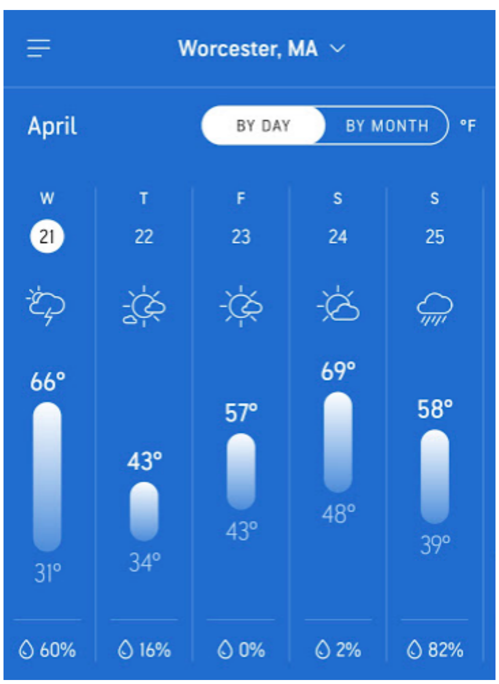
Appendix 1: Preliminary Survey Questions

Section 1: General Weather App Information

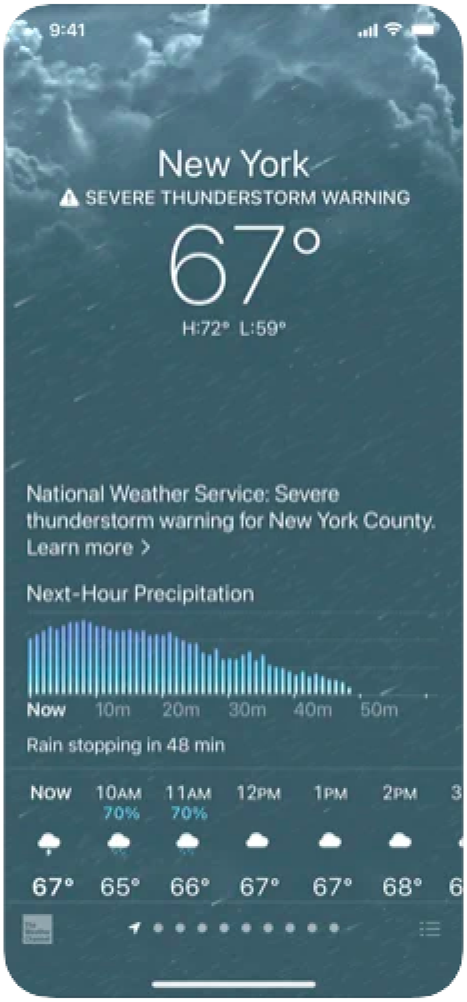
How often do you use a weather app?
What do you use a weather app for?
What is your favorite weather app/website?
What features make this weather application good?
What could be improved in this weather application?
How well do you feel like you understand what a chance of precipitation means?
How accurate do you feel your weather app of choice is in informing you of the chance of precipitation?
How effective do you feel your weather app of choice is in communicating precipitation uncertainty?
How much effort does it take to get a good idea of the chance of precipitation from your weather app of choice?
How reliable do you believe your weather app of choice is for making decisions based on precipitation?
If a visualization of precipitation uncertainty was added to your favorite weather app, how likely would you be to use it?

Section 2: Comparing Weather Apps

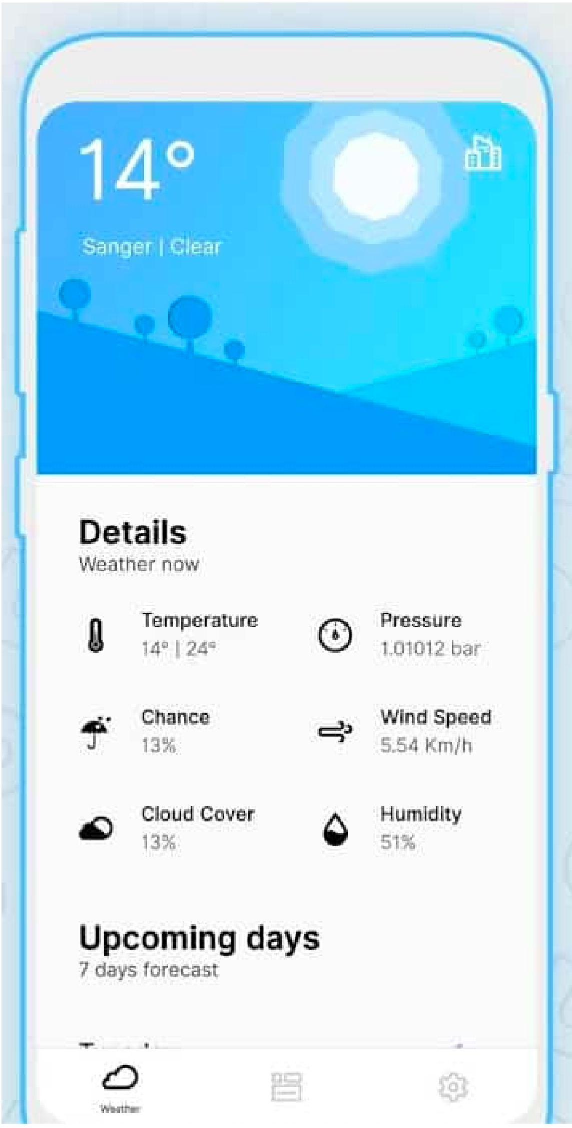
Regardless of what your favorite weather app is, which rain uncertainty visualization(s) from the following app screenshots do you prefer? Why? Are they more effective, easy to read, etc.?
Which ones do you dislike the most, if any?
AccuWeather



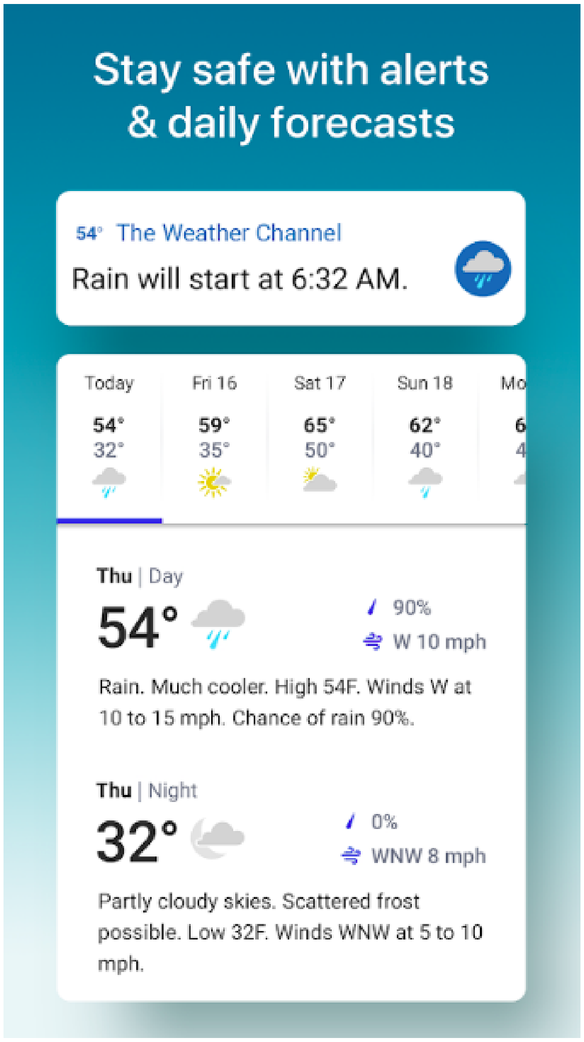
Apple



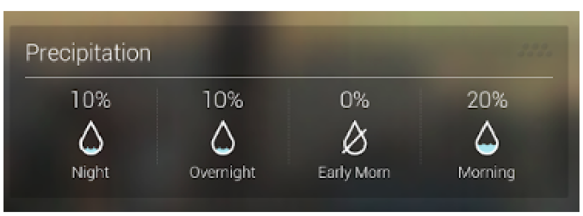
Overdrop



Weather Channel



Yahoo



Section 3: Final Thoughts

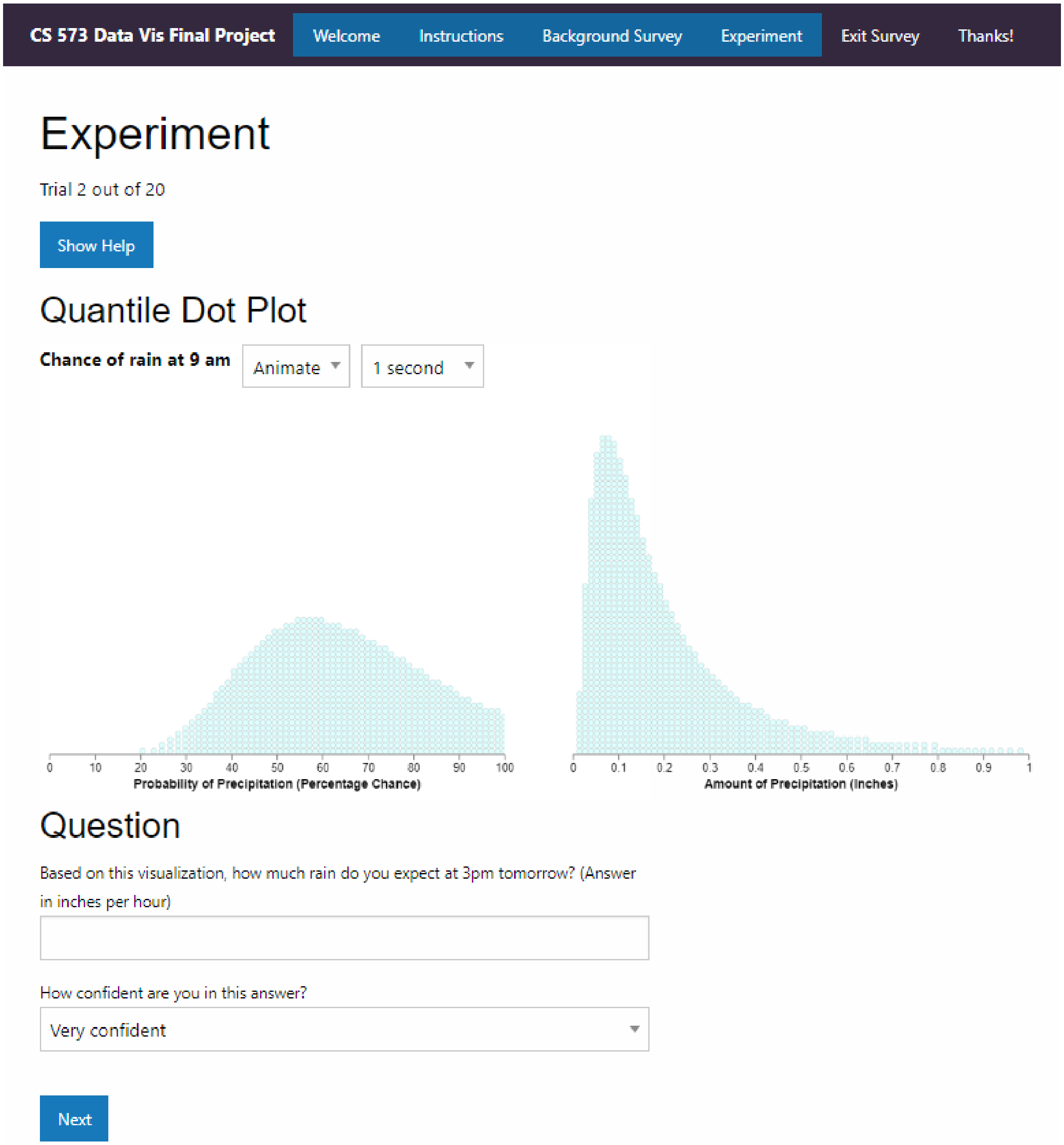
Do you have any other thoughts on improving the display of precipitation uncertainty in a weather app? Any Comments, Questions, Concerns, Feedback?

Appendix 2: Example Survey Response JSON

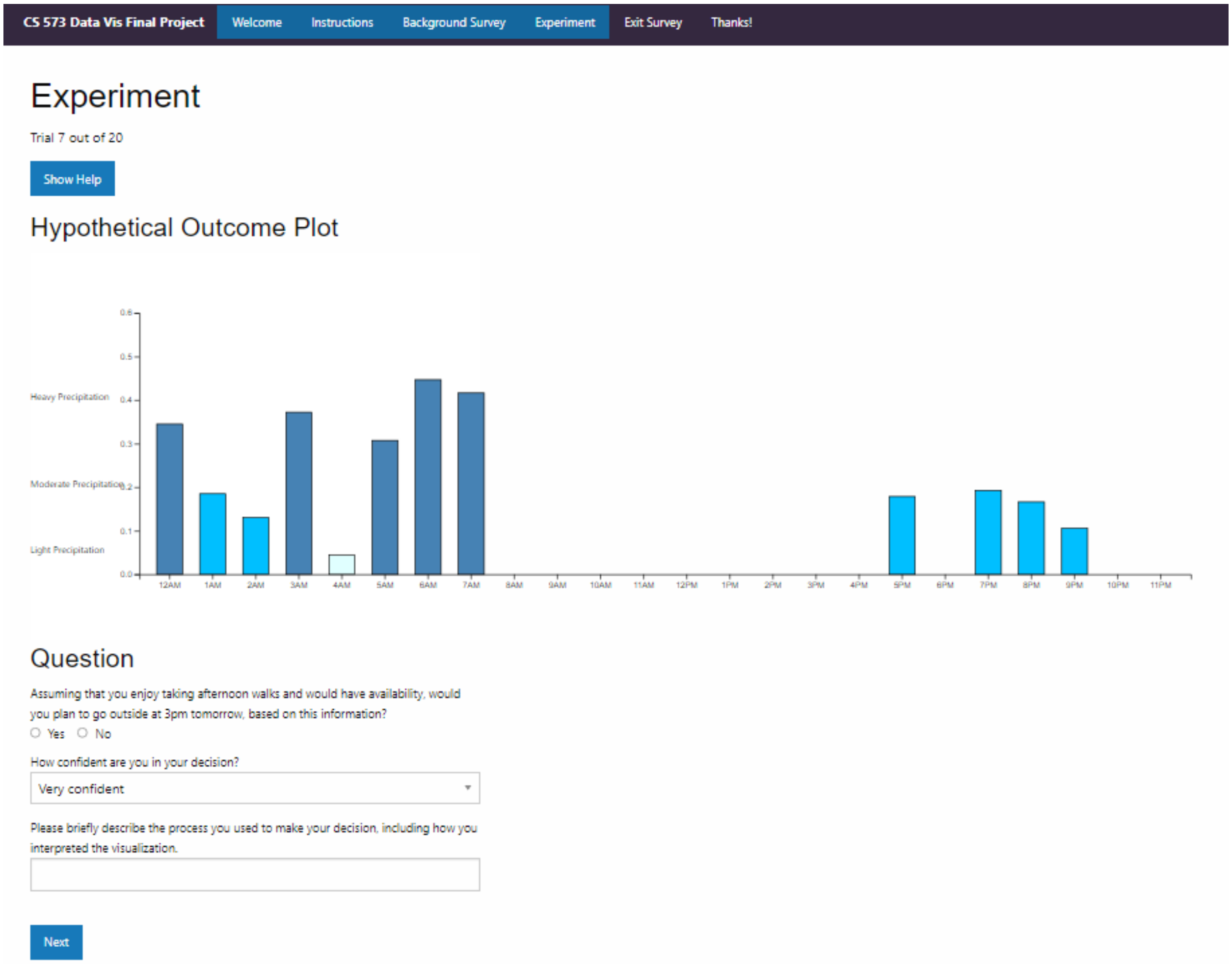

```
{
	Background questions:{
		WeatherAppFrequency: “”,
		FavoriteWeatherApp: “”,
		ExperienceWithApp:””,
		AppReliability:””,
		OftenInaccurate: “”,
		ExperienceWithStats: “”

	},
	Experiment:{
		Trials:[
			{
				HelpButtonHits: 123 ,
				TypeOfVis: “”,
				AmountOfRain: 0.4,
				ChanceOfRain: 0.2,
				QuestionAsked:””,
				WouldYouGoOutAt3pm: “”,
				ConfidenceIn3pm: “”,
				DecisionProcess:””,
				AmountExpectedGuess: 0.21,
				ConfidenceInAmount: “”,
				ChanceOfRainGuess: 0.3,
				ConfidenceInProbability “”
			},
		]
	},
	FollowUpQuestions:{
		VizRank: [“text”,”hop”,”bar”,”dots”],
		Effort:{
			Text:”easy”,
			Bar:”hard”,
			HOP:”Fun”,
			Dots:”what?”
		},
		AdditionalFeedback: “”
	}
}
```

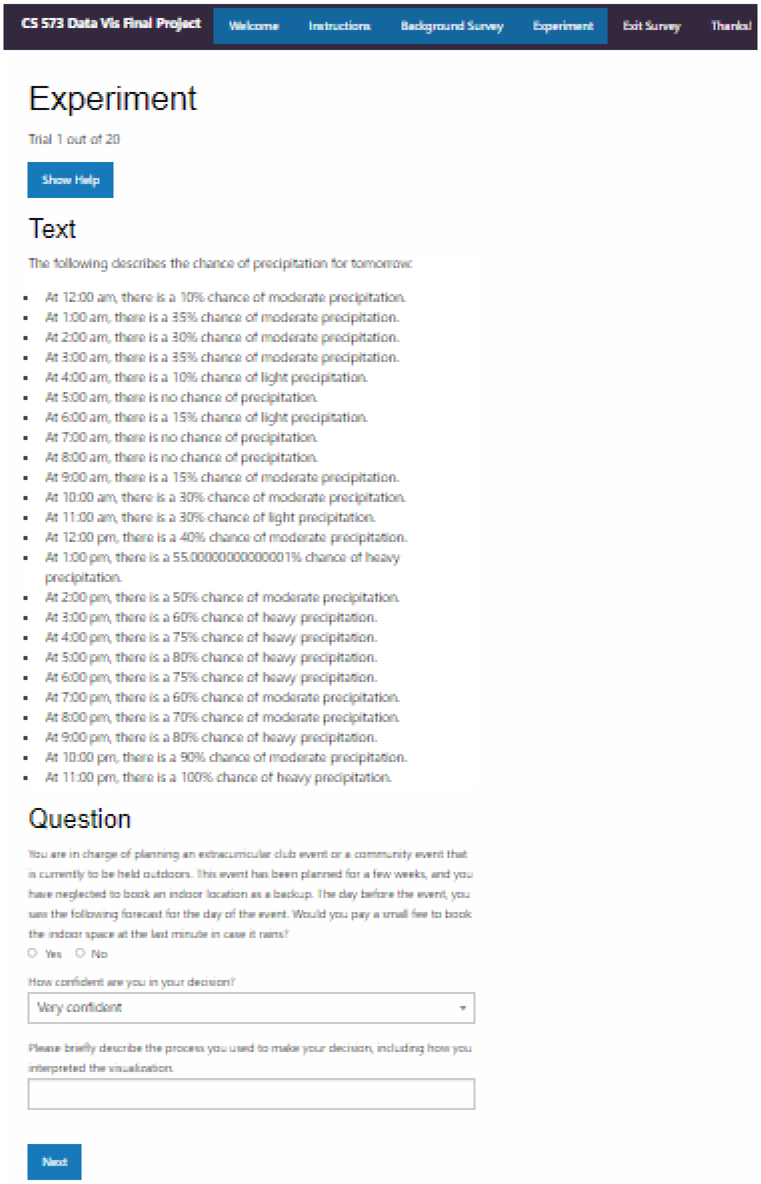
Appendix 3: Example Survey Questions



An example of a vis readability survey question showing a Quantile Dot Plot and the amount of rain question.



An example of a decision making survey question asking users if they would go outside tomorrow afternoon based on the Hypothetical Outcome Plot.



An example of a decision making question asking about planning an outdoor event based on the text based weather data.