Institute for Visualization and Interactive Systems

University of Stuttgart
Universitätsstraße 38
D–70569 Stuttgart

Masterarbeit Nr. MCS-0006

Investigating Objective and Subjective Effects of Aggregating Forecasts from Multiple Sources

Emre Avci

Course of Study:	Computer Science
Examiner:	Prof. Dr. Albrecht Schmidt
Supervisor:	Dipl.-Inf. Miriam Greis, Dr. Tonja Machulla
Commenced:	March 31, 2016
Completed:	September 30, 2016
CR-Classification:	H.5.2

Acknowledgements

I would like to thank everyone that helped and supported me during the creation of this master thesis. A very special thanks goes to my girlfriend, Nazli, who was always there for me when I needed and who is the sole reason I was able to get to this point in my life and finish this master thesis. Also, I would like to thank my supervisors Dipl.-Inf. Miriam Greis and Dr. Tonja Machulla for their useful comments, remarks and support through the learning process of this master thesis. Further thanks goes to the participants in my survey and study, who have willingly shared their precious time during this process.

Abstract

Uncertain data is found in many different contexts in life and often people need to make decisions based on this uncertain data. An example of such uncertain data is weather forecasts. People make decisions of varying importance, based on the weather forecast, almost everyday. In order to increase their confidence and make an informed decision, many people compare weather forecasts from multiple sources, which is a very time consuming and cumbersome task. This master thesis aims to find the best method to combine multiple weather forecast sources in a single application, so that the users would have the confidence of comparing multiple sources without the time consuming nature of doing it manually. We have identified three methods to aggregate weather data from multiple sources and compared them using different representations. Together with our aggregation methods we have also included a single source option in our research to compare aggregated weather forecasts with single source forecasts. We have conducted a survey to decide on which aggregation method to test further in our research and then implemented a mobile application to test our aggregation method in real world user study. Our results showed that independent of the representation method, users prefer aggregated weather forecasts over single source forecasts. Additionally we have found that users prefer aggregations that provide more information when making important decisions and prefer computationally aggregated methods for daily use.

Contents

List of Figures

List of Tables

1 Introduction

Weather is involved in every aspect of life and most of the time its effect is significant on people's decisions. Humans have been forecasting the weather since the early civilizations [Cen05]. The art of weather forecasting has improved greatly over the years and has turned into a science after the invention of measurement instruments after Renaissance [Cen05]. Despite all the progress in this field, weather forecasts are still very uncertain, especially when investigated for a future date. As a result of this uncertainty, different sources can have completely different forecasts depending on their own measurements. This variety of forecasts from different sources, cause people to lose their trust on weather forecasts. The way people deal with this uncertainty of weather forecasts vary. Some people choose to prepare for the worst possible case where others check multiple forecasts and compare them to make a decision.

With the increasing accessibility of information, people's expectations to get the correct weather information easily is increasing. Yet if one wants to check and compare multiple weather forecasts to get a general sense of the weather, one needs to use multiple applications/websites and find a way to either put the information side to side in order to compare, or keep the values from each source in mind to compare with the next one. This is both time consuming and complex task. And the efficiency of this kind of comparison depends highly on the skill of the user to interpret and understand multiple uncertain data coming from different sources. The user should find a strategy to combine this information and it is possible that users might have non-optimal aggregation techniques.

In order to make this approach of comparing and aggregating multiple weather forecast sources easy to use and accessible for everyone, an optimal method for aggregating multiple weather forecast sources is required. In this thesis, we compare four different methods to aggregate weather information from multiple weather forecast sources in terms of usability, user confidence and preference with consideration of the representation method for the weather data and the scenarios the weather data is used in. The most successful method in terms of these measures is implemented as an Android application, and a user study to see the effectiveness of the application is done using this application.

1.1 Scope of this thesis

This work focuses on development and evaluation of aggregation methods for combining multiple weather forecast sources in one application. This includes the development and evaluation of possible representation methods for weather data, which may potentially affect the effectiveness of an aggregation method. Moreover, both the aggregation and representation methods should be evaluated considering possible scenarios of varying importance, in order to include the effect of events and their importance to users' preferences. For each combination of aggregation and representation methods, we have developed a static prototype to be used in the evaluation. We have conducted an online survey to determine the best combination of aggregation and representation method to include in the Android application we made. Later a user study has been conducted to evaluate the aggregation method under real life circumstances and get feedback from participants to improve the application and the aggregation method. The analysis and discussion of the results completes the assignment of this thesis.

1.2 Outline

Following this, we start looking at the related work in **Chapter 2 - Related Work**, where we discuss the existing literature on uncertainty, weather forecasts and visualization from researches in the fields psychology, HCI and visualization. Then in **Chapter 3 - Sketches**, we present the aggregation methods and representation methods that we have came up with and explain the motivation behind our decisions as well as the sketches we have created based on these methods for our online survey. The online survey is explained in detail in **Chapter 4 - Online Survey**, where we also talk about the composition of the survey, participants as well as the results and our findings from these results. Based on the requirements we have gathered from the results of the online survey, we have built an Android application which we discuss the details in terms of design and development in **Chapter 5 - Android Application**. In order to see how our methods perform in real life conditions, we have used the Android application we have built to run a user study. The details about how we conducted and evaluated the user study is discussed in **Chapter 6 - User Study**. In this chapter we also present the results of the user study and our conclusions based on these results. The final conclusions are presented in **Chapter 7 - Summary and Conclusions**, where we also talk about possible future improvements and research topics that could emerge from our work.

2 Related Work

In this chapter, we present background information, relevant literature, and other sources that covers the basis of our work. Our work builds up on previous work from various fields including, psychology, HCI, visualization and meteorology and the major topics are uncertainty, weather forecasts and uncertainty visualization.

2.1 Uncertainty

Uncertainty is used in many different fields with varying meanings [MRH+05]. The understanding of uncertainty differs according to the domain, the term is used in. Douglas Hubbard defines uncertainty as follows:

"The lack of certainty. A state of having limited knowledge where it is impossible to exactly describe the existing state, a future outcome, or more than one possible outcome." [Hub10]

There are definitions that only work in the context of a single domain, for example for international relations [Rat07], for water management [Bec87] and for economics [Bra16]. W. E. Walker et al. provide a general definition of uncertainty, based on the common parts of the definition of uncertainty from different fields, as the following:

"Uncertainty is any deviation from the unachievable ideal of completely deterministic knowledge of the relevant system" [WHR+03]

Further, W. E. Walker et al. discriminate the uncertainty to three dimensions, location, level and nature of uncertainty and uses the existing typologies to explain and support these dimensions [WHR+03].

2.2 Uncertainty in Weather Forecasts

Weather forecasts are uncertain by their nature, yet the information about weather forecast uncertainty is not presented to the most forecast users [MDL08] [Pal00] [SP11].

Presenting the uncertainty information in weather forecasts still remains as a challenge since the people's understanding of uncertainty information varies [MDL08]. Studies showed that people prefer forecasts with uncertain information compared to deterministic forecasts and they make better decision when using a forecast with uncertain information [JPJ+07] [NJ09] [RBKS06] [RK09]. Despite this, several surveys conducted by researchers show that most of the people do not interpret the uncertain information correctly [GRM+08] [Nic99]. There are existing work on understanding public's perspective over the uncertainty of weather forecast, which shows that more than half of the people does not understand the information provided by uncertainty information in weather forecasts correctly [MDL08] [RBKS06] [HS13] [HP07]

2.3 Uncertainty Visualization

Visualization techniques mostly deal with data that is free from uncertainty, yet recently the visualization community started to work on the challenge of uncertainty visualization more frequently [BAL12]. Sanyal et al. [SZB+09] evaluates the most common techniques for visualizing uncertain data for 1D and 2D data sets. These common techniques are the following:

- Scaling the size of glyphs
- Altering the color attribute of glyphs
- Color-mapping the surface of data with uncertainty
- Traditional errorbars

They have found that traditional errorbars consistently underperformed where the efficiency of other techniques were task dependent [SZB+09].

The work in uncertainty visualization spans over various domains. MacEachren et al. [Mac92] evaluates the usage of Bertin's basic system of graphic variables (location, size, value, texture, color, orientation, shape) and two additional variables (color saturation and focus) for visualizing uncertainty in cartographic visualizations. Lipkus et al. [LH99] evaluates uncertainty visualization in the context of communicating risk. Hoffrage et al. [HLHG00] looks at uncertainty visualization for statistical information.

Olston et al. [OM] introduces a new visualization technique called *Ambiguation* that works on data with bounded uncertainty. They also describe how this technique can be applied to common visualization techniques like line graphs and bar graphs. They have found that *Ambiguation* performs better when compared to errorbars and some

similar techniques. Thus, we have used this visualization technique in our research for representing the range aggregations with line and bar graph representation techniques.

In the context of weather forecasts, Demeritt et al. [PST+13] discusses about the various ways to visualize uncertain weather forecast data and also compares the existing methods used by some forecast providers.

Along with different visualization techniques, there exists some research on how to evaluate and compare uncertainty visualization techniques [BAL12] [SZB+09] [RBKS06] [OM]. Some of the studies especially focus on non-technical people since uncertainty information mostly requires basic statistics knowledge [IM87] [PST+13] [MLD10].

2.4 Measuring Confidence

In order to see how effective a method is, we need a way to evaluate the user's confidence for a given aggregation method. There is not too much literature about measuring confidence specifically thus, we have examined papers that focus on measuring risk, trust and honesty.

Measuring trust is not a trivial work, since simply asking a person would not give you the correct information about whether they trust something/someone or not [Rai00] [YY94]. Thus, there are several approaches like trying to measure trust by using games [PST+13] [GLSS00] [YY94]. When participating in a game, the behaviours of a participant would show whether they trust something or not.

When confidence needs to be measured using a traditional questionnaire, then usage of Likert items are suggested [YY94].

3 Sketches

In this chapter we will explain the process we have followed when developing the methods and the sketches we have used in our research. We will start with the 3.1 - Competitor Analysis where we investigate existing similar applications. Then in 3.2 - Aggregation Methods we will explain the methods we have came up with in order to aggregate multiple weather forecasts, based on our literature research and competitor analysis. In 3.3 - Representation Methods we will explain the representation techniques we have included in combination with the aggregation methods. Here we will also present the sketches we have used in the online survey for each aggregation - representation combination. In 3.4 - Standardizing sketches for the study we will explain how we provided a standard for the sketches to avoid effects of other factors in our online survey.

3.1 Competitor Analysis

Before starting to develop our own methods to aggregate and visualize multiple weather forecast providers, we have researched existing weather forecast applications that make use of information from multiple sources in order to get an initial idea of methods that are already tested with real users.

During our research, we have found two applications that allow comparing multiple weather forecast sources, Climendo and WeatherXM.

3.1.1 Climendo

Climendo[1] is a mobile application that is available in both Apple App Store and Google Play Store. It provides the information from 9 different weather forecast providers (Weather Underground, AccuWeather, World Weather Online, Dark Sky, SMHI, YR, National Weather Service and Foreca). The number and name of the sources vary

[1] **Climendo:** http://climendo.com/

Figure 3.1: User interface of Climendo

depending on the location of the user. They calculate an accuracy value for each weather forecast provider using historical data for user's location and only display those sources with good accuracy to the user. Then they also calculate an average weather forecast, using the data from these providers, to show the user a simpler forecast information. The weather forecast calculated by Climendo itself is simply the average of all values from the weather providers, but they also provide an indicator that shows how much the different sources agree, which they call the certainty of the weather forecast.

The visualization of Climendo uses piktogram images to display the weather information. They normally only show the average calculated weather forecast and the certainty indicator, yet it is possible to expand it to view all the individual weather forecasts from different sources for direct comparison. The interface of Climendo can be seen on Figure 3.1.

3.1.2 WeatherXM

WeatherXM[2] is an Android application that compares multiple third party weather forecasts for user's location and ranks them according to their past performance using historical data. Then it displays all the sources to the user, sorted by accuracy rating. This way the user can see all the sources together with their accuracy ratings side by side and can decide on which weather forecast provider to trust. WeatherXM also provides some graphs that show the accuracy of weather forecast providers over time, if the user wants to learn how consistent the weather forecast provider has been.

The visualization of WeatherXM uses piktogram images. They normally show the piktogram and the minimum/maximum temperature of the day, but users can click on a particular source to see hourly values for temperature, chance of rain and wind.

WeatherXM does not provide an aggregated weather forecast, instead they leave the aggregation to the users and just display all the sources next to each other for easy comparison. The interface of WeatherXM can be seen on Figure 3.2.

3.2 Aggregation Methods

In order to best utilize multiple weather forecast sources, an optimal aggregation method should be selected. In order to understand which aggregation method works best, we have picked four different methods to evaluate in the survey, Single Source, Mean, Range and Direct Comparison. A simple illustration of aggregation methods is presented in Figure 3.3 The aggregation methods are chosen in a way that each method requires different amount of work from the user's side.

3.2.1 Single Source

Though not actually an aggregation method, displaying only a single source is included in the survey in order to compare the aggregations with non-aggregated weather information. Since this is the most common way of providing weather information, it also allows us to compare our methods with the existing applications. In this method we only use one source and provide the information as it is.

[2] **WeatherXM:** http://weatherxm.exm.gr/

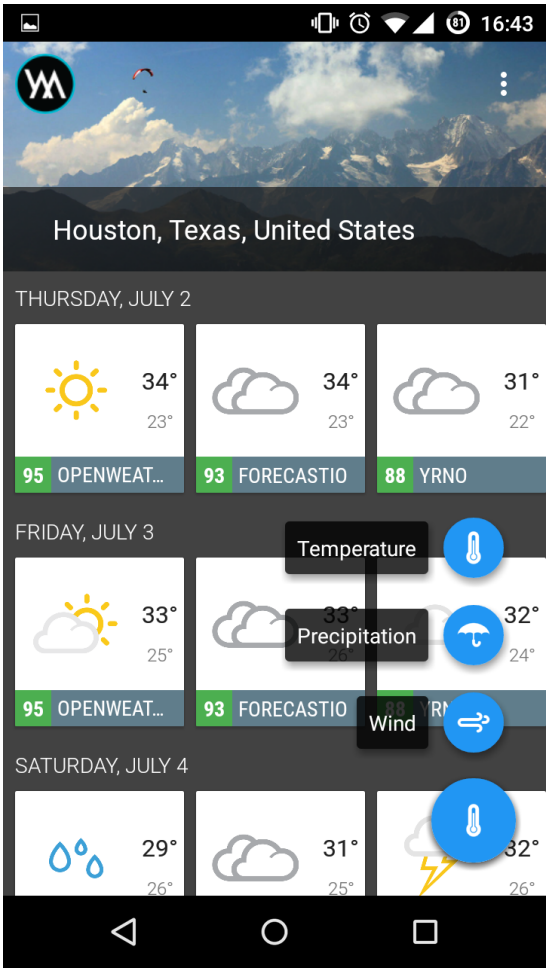


Figure 3.2: User interface of WeatherXM

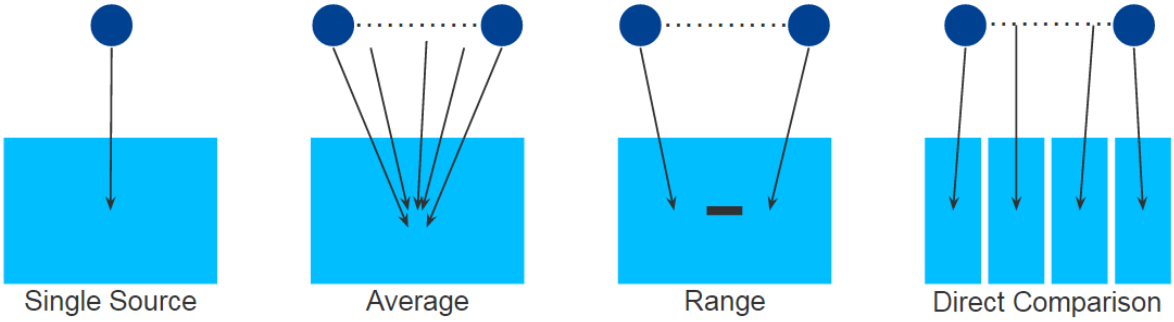


Figure 3.3: Aggregation methods

3.2.2 Mean

This aggregation method simply provides the mean of the values from different sources. It acts as a quick way to reduce multiple values into a single, easy to interpret value. This method keeps the simplicity of having a single source while also providing the additional information coming from multiple sources. The mean can also be extended to work as a weighted mean by using the accuracy ratings of sources. Accuracy ratings can be calculated using historical weather data to check how accurate the weather forecast source has been over a certain period of time. When the mean aggregation method is used, it is important to communicate this properly to the user, since without additional information it is not possible to distinguish mean aggregation method from single source. Possible ways would be to show the names of the sources, provide an option to see individual sources or make sure the interface copy communicates that the value is calculated using multiple sources.

3.2.3 Range

This aggregation method provides a range that spans over the values coming from multiple sources. It shows the minimum and the maximum possible value. Though this aggregation method provides more information, it also requires more work from the user's side to be interpreted correctly. In addition, having one source with an extreme erroneous value would affect the outcome of this aggregation method significantly.

Along with the information about the possible values the weather may have, the range aggregation method also provides information about how much the different sources agree. When the range is wide, the user understands that the sources differ greatly, when the range is very narrow the user would know all the sources agree on the same value.

3.2.4 Direct Comparison

Direct comparison is the aggregation method that requires the largest amount of work from the user's side. It provides the information from each individual source independently so that the users can aggregate the information by themselves. The values from different sources should be visible next to each other and they should be presented in a way to allow easy comparison of similar values. Direct comparison method requires the highest amount of effort from the user since the users need to make their own judgement about the weather by comparing, aggregating or interpreting the raw data from multiple sources. The effectiveness of this aggregation method depends on the

number of weather forecast sources used, since comparing more than three sources would be very difficult and time consuming for most of the users.

3.3 Representation Methods

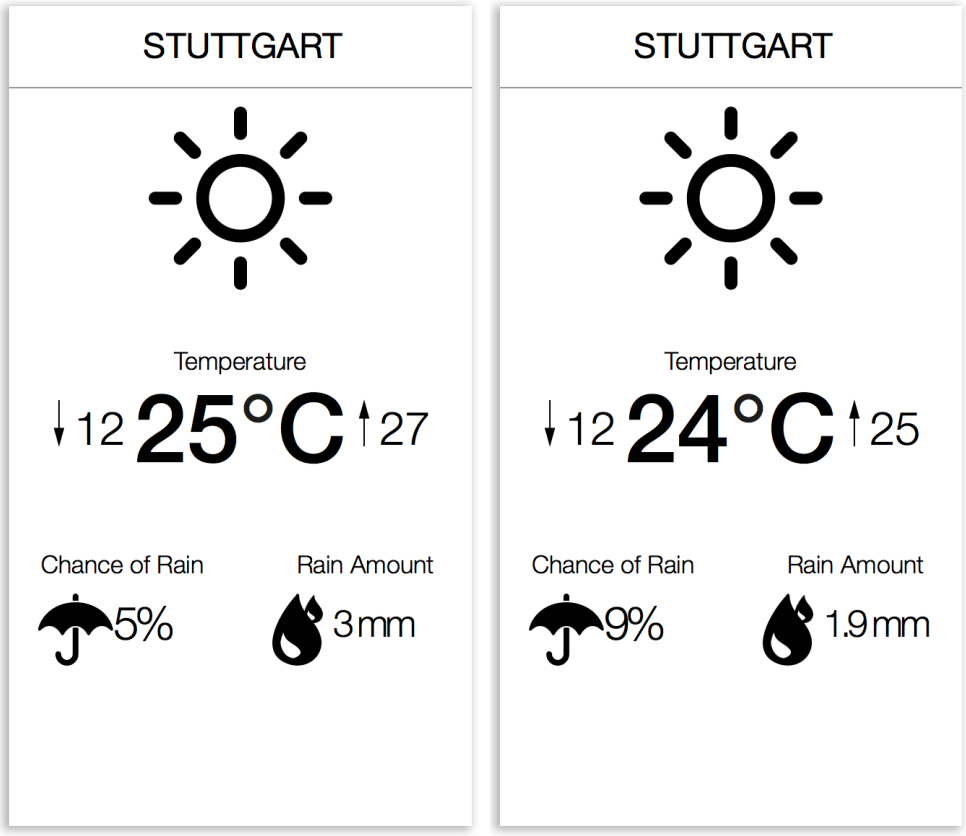
Humans have a high ability to recognize and remember graphics [Sta73], thus the representation technique plays an important role in communicating data to users. Since the aim of our research is to find the best way to combine weather forecast data from multiple sources, how we represent our aggregated data may impact the effectiveness. Thus, we have chosen four different representation methods in order to consider the influence of representation method on the effectiveness of an aggregation method in our questionnaire. We have used the same type of visualization for temperature, amount of rain and chance of rain on each evaluated screen. For some of the aggregation methods, we have slightly modified the visualization techniques to match the aggregate data better.

3.3.1 Pictogram Visualization

It is important for a weather application to communicate the information as simple and quick as possible. Pictograms are the most commonly used visual factor when representing weather data. They have been used to represent weather data in a variety of mediums for years and people are very familiar with the meanings of different weather pictograms. Thus, the use of pictograms as the representation method increases the learnability of the application and makes it easy to onboard new users, since they do not have to spend the time and effort to learn how the representation works.

In order to make sure that people's familiarity with the representation technique is not the deciding factor in our survey, we have decided to include a pictogram image in all of our other representation techniques as a side element. Yet, we still kept pictogram representation as a separate representation method for comparison. In this representation we used the pictogram as the main element in the screen.

For each of our aggregation methods, we have kept the pictogram without any modification since modifying or combining multiple pictograms loses the purpose of using a pictogram due to it no longer being simple and easy to understand. When aggregating multiple weather forecast sources into one representation, we have used the pictogram image that the majority of the sources suggested.



(a) Single Source **(b)** Mean

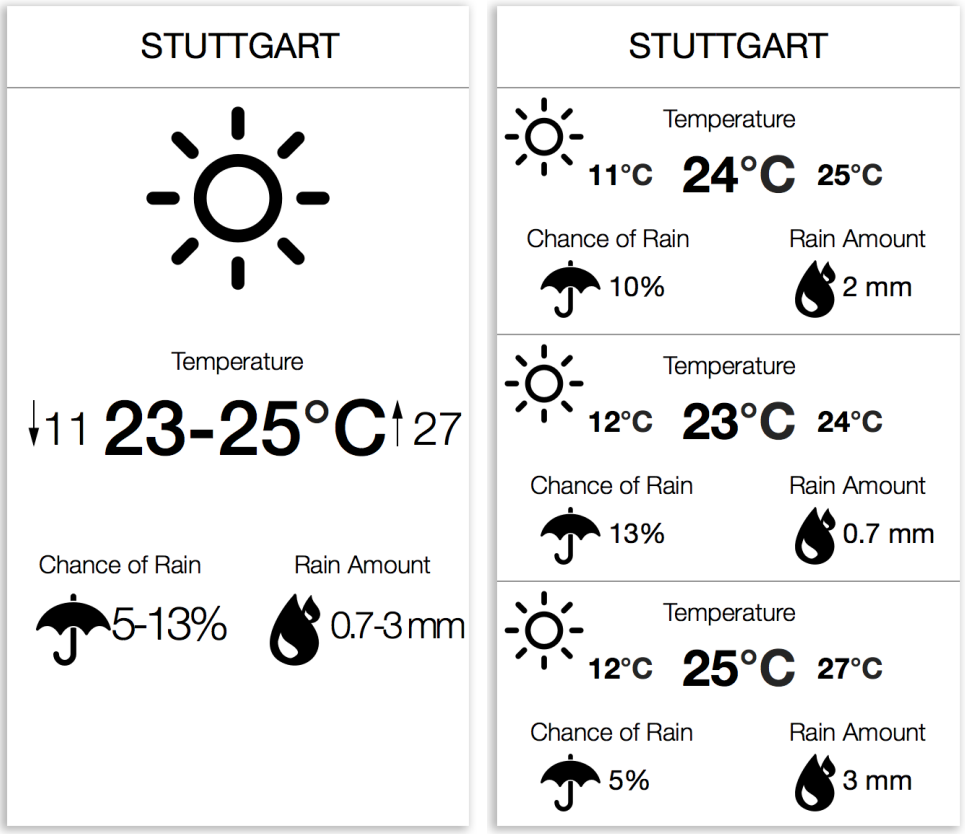
Figure 3.4: Sketches for Pictogram representation method (1)

The sketches we have created for each aggregation method using pictogram representation is shown in Figure 3.4 and Figure 3.5.

3.3.2 Bar Graph Visualization

Bar graphs are very common for comparing quantitative data, due to human's increased ability to compare lengths of objects compared to other visual properties. Since our application mainly aims at making it easier to compare multiple weather forecast sources, we wanted to measure the effectiveness of bar graphs in this context.

Bar graph visualization is most suitable for direct comparison aggregation method, since we have multiple bars next to each other for easier comparison, and this scenario fits the purpose of bar graph visualization best. For mean and single source aggregation methods, where there is a single aggregated value, we have modified the bar graph visualization slightly to benefit from the usage of length visual property here as well.

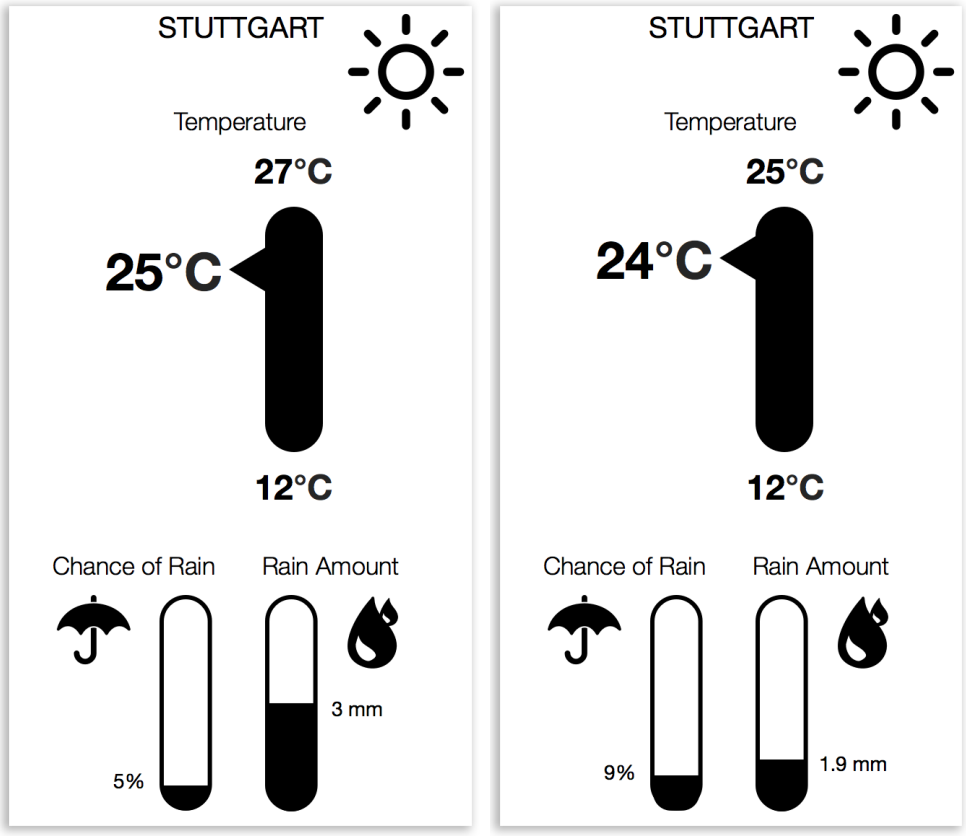


(a) Range **(b)** Direct Comparison

Figure 3.5: Sketches for Pictogram representation method (2)

We have created the bar graph by setting the minimum and the maximum temperature to be the lowest and highest points in the bar. The length of the bar communicates the difference between lowest and highest temperatures of the day. In addition we have marked the point where the average temperature of the day is, so that it is possible to see whether it is closer to the minimum or the maximum temperature in average throughout the day, by comparing to distance from the marked point to both ends.

When using the range aggregation method, the data from multiple sources create a bounded uncertainty. In order to represent this uncertainty using bar graphs which is shown in 3.7a, we have used the Ambiguation method developed by Olston et al. [OM]. Which uses a different shade in a section of the bar to represent the uncertain part of the bar graph. This way the bar would be one color until the point where the data is certain, which also marks the minimum possible value for that data and the bar would continue with a different color up till the point where the maximum possible value is, and the different colored part would show that the value would be anything between that area.



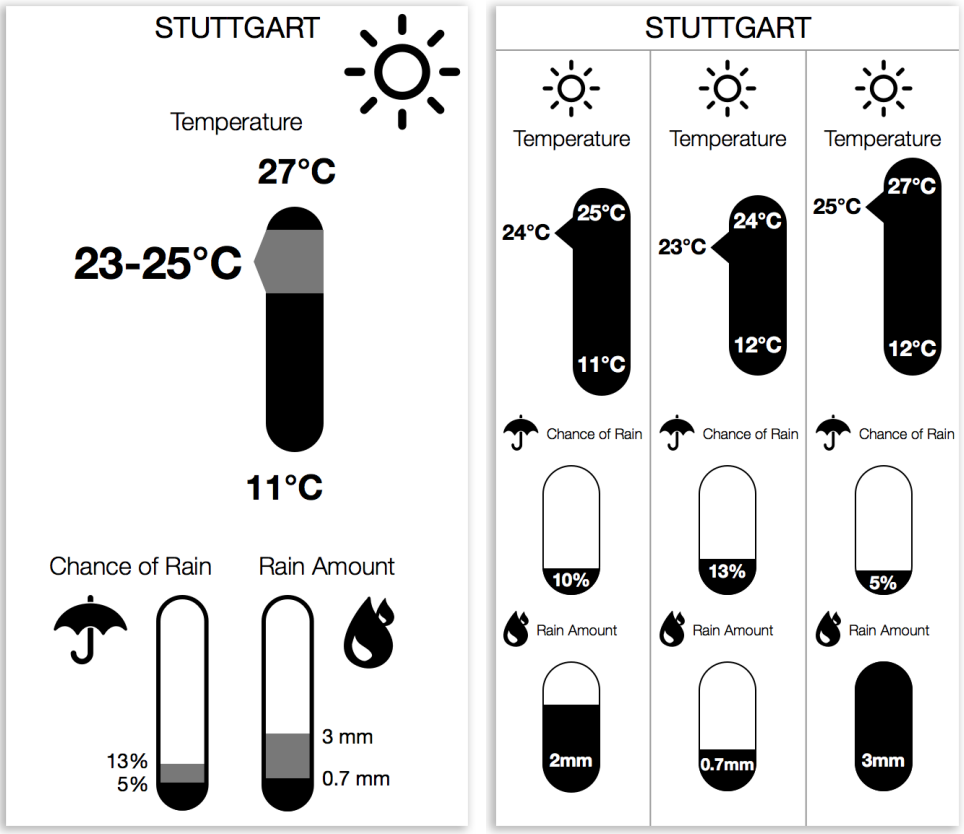
(a) Single Source **(b)** Mean

Figure 3.6: Sketches for Bar graph representation method (1)

The sketches we have created for each aggregation method using bar graph representation is shown in Figure 3.6 and Figure 3.7

3.3.3 Line Graph Visualization

Line graphs are often used to represent temporal data in a timeline like representation. Since the weather data is temporal, we have also evaluated line graph visualization in our survey. We have plotted the values for temperature, chance of rain and amount of rain over the hours in a day to make it easy to have an overview of the day. We have created three different graphs for temperature, chance of rain and the amount of rain and, we have used the same horizontal axis for each line graph, so that user can easily see the information for a certain hour of the day in all three graphs simultaneously. For the direct comparison aggregation method, we have plotted multiple lines in the same graph and labeled them with different colors to make comparison easy. Line graphs used



(a) Range **(b)** Direct Comparison

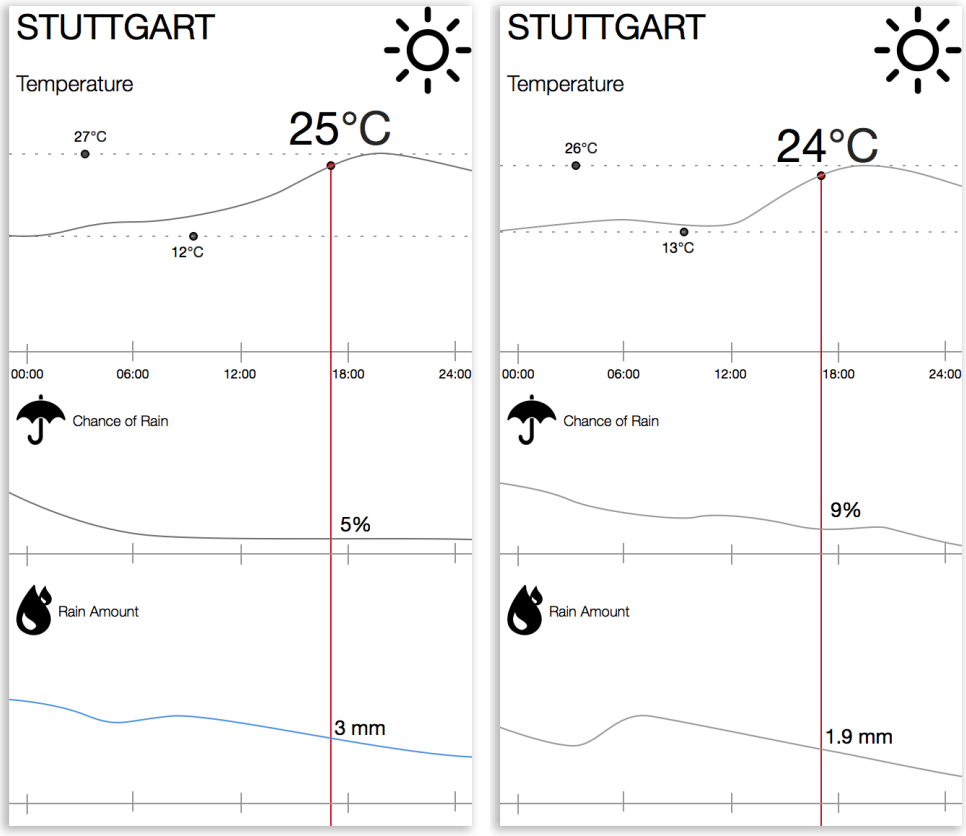
Figure 3.7: Sketches for Bar graph representation method (2)

with the range aggregation method were again created using the Ambiguation method developed by Olston et. al. [OM]. This way by looking at the thickness of the line at a certain point in time, it is possible to understand how much does the sources agree on one value.

The sketches we have created for each aggregation method using line graph representation is shown in Figure 3.8 and Figure 3.9

3.3.4 Text Visualization

Text visualization is a representation method aimed to replicate the feeling people get when listening to weather forecast on radio or television. It represents the data embedded in a block of text explaining the weather of the day. The values are highlighted with color to enable quick glance and to let power users consume the information in a much shorter time. For the direct comparison aggregation method, we have put



(a) Single Source **(b)** Mean

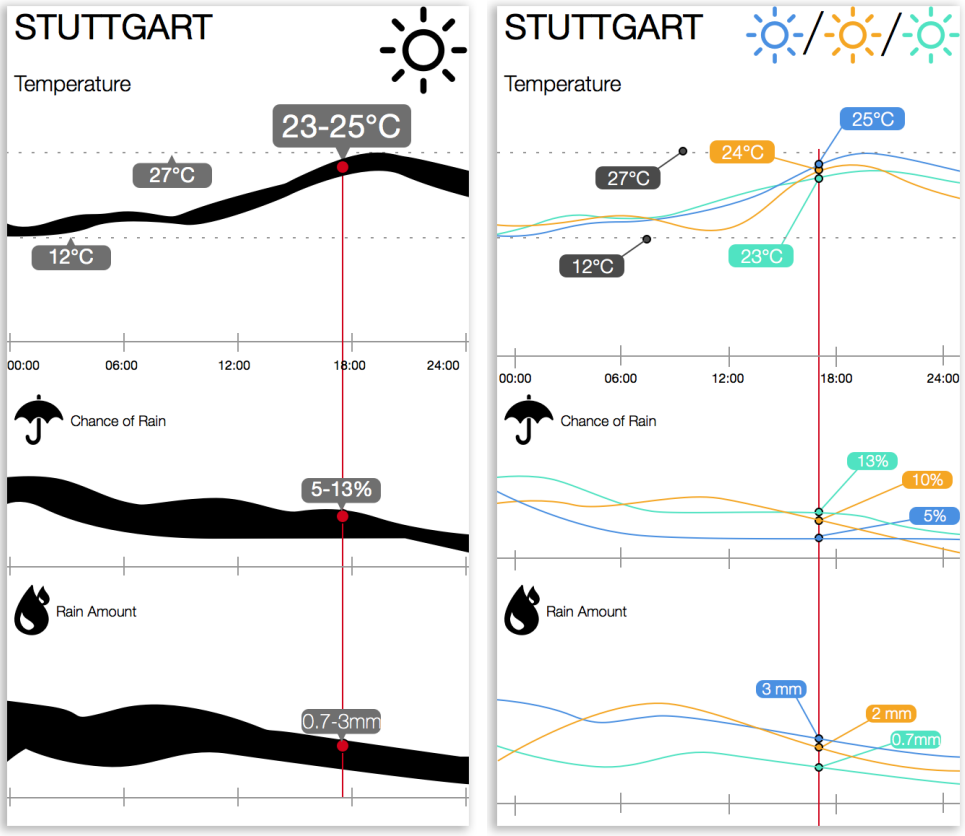
Figure 3.8: Sketches for Line graph representation method (1)

the values next to each other and separated them with a forward slash. For range aggregation we have presented the values in the form 'x to y' in order to make it sound more natural.

The sketches we have created for each aggregation method using text visualization representation is shown in Figure 3.10 and Figure 3.11

3.4 Standardizing sketches for the study

In order to focus the survey on comparing only the aggregation methods, we have decided to standardize every other aspect of the sketches. Since our representation methods also create a great difference in understanding of different aggregation methods, we kept the standardization in the representation method level. And together with this we have also structured the survey in a way that only the aggregation methods for one



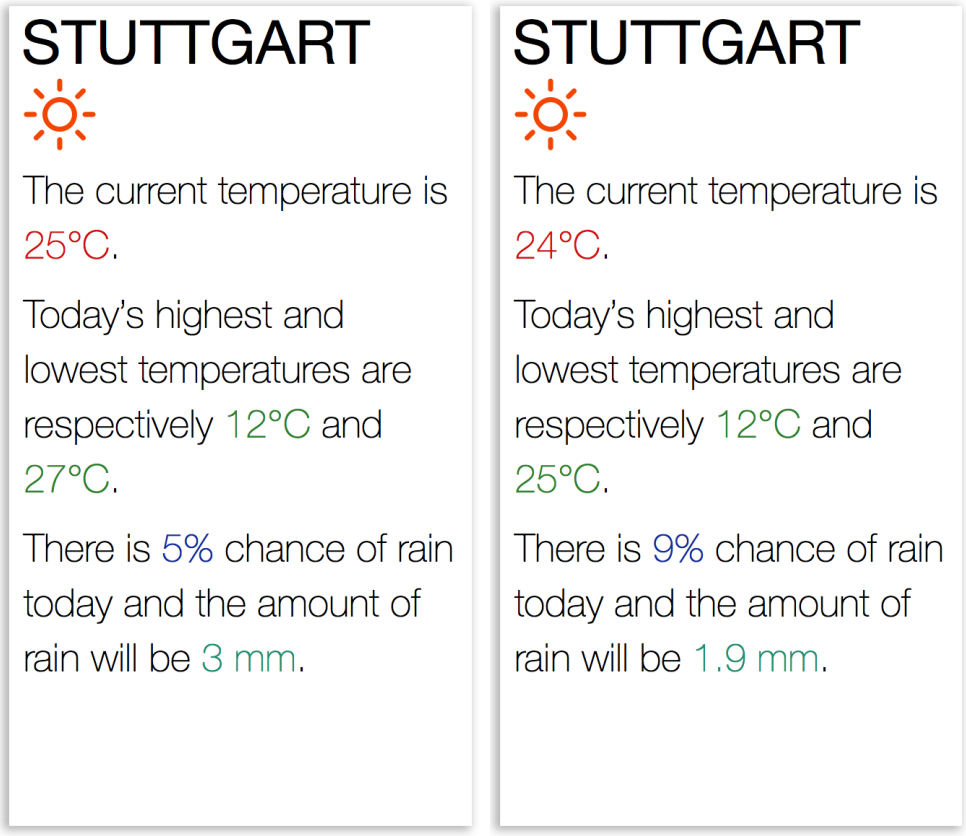
(a) Range **(b)** Direct Comparison

Figure 3.9: Sketches for Line graph representation method (2)

representation method are compared at a time. This way we have avoided interactions between a representation method and an aggregation method on the final results.

As part of the standardization process, we have decided to minimize the amount of information presented in each sketch. We have used the results of a small survey we have conducted about the importance of weather data, which provided us with the information about the most important data in a weather forecast. We have decided to have the following rules for standardization of each sketch:

- Each sketch should have a pictogram
- Each sketch should represent the current temperature, minimum/maximum temperature, chance of rain and amount of rain, since these are the most important weather data elements according to our survey.

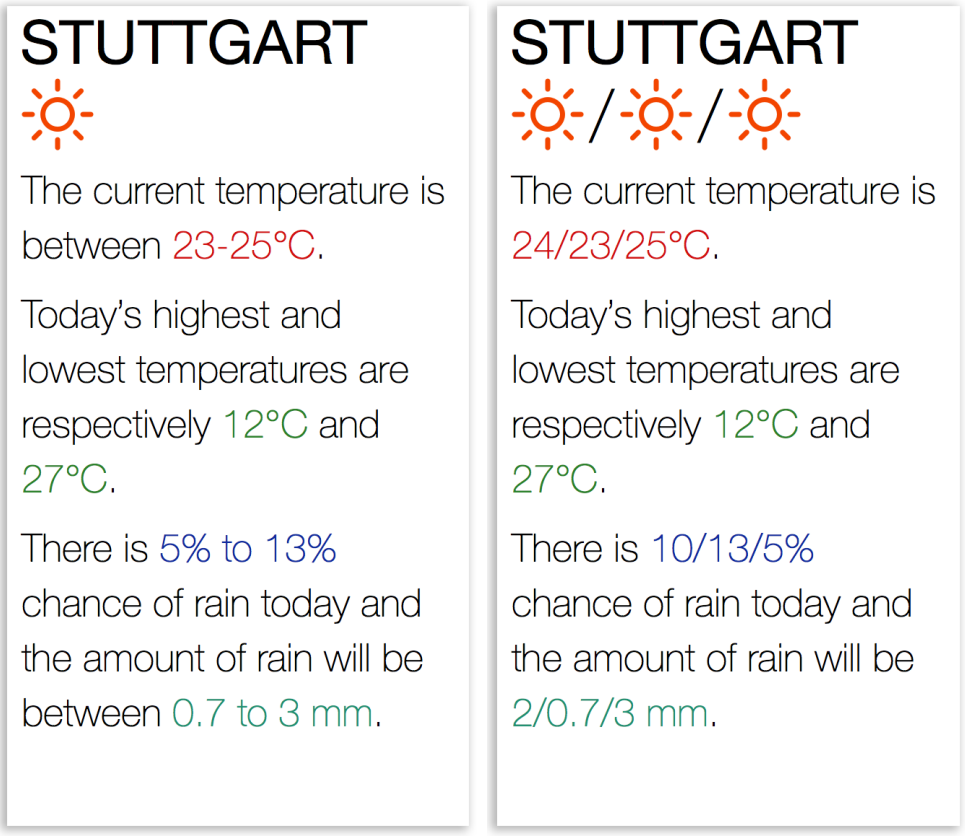


(a) Single Source **(b)** Mean

Figure 3.10: Sketches for Text visualization representation method (1)

- Each sketch should use the same visualization technique for visualizing temperature, chance of rain and amount of rain, according to the visualization technique it represents.
- Each sketch should display the name of the location which the weather data is for.
- The layouts of sketches of one visualization technique should be as similar as possible, to make aggregations easy to compare.

The standardization process made sure that for one visualization technique, the only difference between the sketches were the aggregation type, thus we were able to compare the aggregation methods without the influence of other factors in the design.



(a) Range **(b)** Direct Comparison

Figure 3.11: Sketches for Text visualization representation method (2)

4 Online Survey

The aggregation and visualization methods we have prepared needed a user evaluation in order to see which aggregation-visualization combination was preferred by the users most. And also to see whether the users would like to use multiple weather forecast sources in their daily lives or not. The major goal of the online survey was to find out which aggregation method works best in combining multiple weather forecast sources.

4.1 Survey Composition

When we were developing the aggregation methods, we have realized that the effectiveness of an aggregation method may depend on the visualization method used to represent the aggregated information. In addition each aggregation method requires different amount of work from the user's side, for example using direct comparison, the user needs to take the time and effort to compare each different source whereas in the mean aggregation, the aggregation method already does that for the user and user can just use it as if that was a single source. People's will to spend time and effort in understanding the weather depends on how critical the information is for them. In order to make sure we have considered all possible factors that can affect the preference for an aggregation method, we have developed three different scenarios. These scenarios have varying importance, which we have decided by running a small survey among 15 participants, we have presented them with 12 different scenarios we have came up with and asked the to rate these scenarios from least important to most important. Using the results of this survey, we have decided on the following three scenarios in increasing importance:

Daily Dress Choice: You are going to work/school in the morning on a regular day and you need to decide what to wear.

Outdoor BBQ party: You are planning a BBQ party for tomorrow and you want to know whether the weather will be good for having a party outdoors.

Outdoor Wedding: You have your wedding tomorrow and it is an outdoor wedding. You need to decide whether you should make changes in the organization (like arranging a tent, or moving the wedding indoors).

In the next step, we have compared the aggregation methods for each visualization method and for each scenarios. Together with this, we have also prepared questions for demographics and general information about the weather forecast usage of the participants. We have asked questions to understand how often do people check weather forecast, what sources are they using, whether people are already trying to compare multiple sources by themselves, do they trust in weather forecast etc. The list of all questions in the survey is provided in Appendix A.

4.2 Mini Survey for Scenarios

The importance of weather data depends on how much impact weather events have over the user. And according to this importance, user's willingness to spend time and effort to learn about the weather changes. Simultaneously this also affects the expectations of users from a mobile weather application. In order to measure the effect of this aspect on participants' choice of aggregation method, we have decided to include scenarios of varying importance as a factor in our survey. We have presented participants with three scenarios, one casual event, one somewhat important event and one very important event, and we have asked them to evaluate each aggregation-representation method combination separately considering these scenarios.

In order to make sure that the chosen scenarios have the intended importance level, we have decided to conduct a hallway questionnaire, where we asked 15 participants to rank the importance of 12 scenarios we have come up with. Participants marked each scenario with one of casual, somewhat important and very important ranks, and we have used the scenarios that were selected by most participants in each importance level, in our survey. The selected scenarios are the following, in order of increasing importance:

- **Daily Dress Choice:** You are going to work/school in the morning on a regular day and you need to decide what to wear.
- **Outdoor BBQ party:** You are planning a BBQ party for tomorrow and you want to know whether the weather will be good for having a party outdoors.
- **Outdoor Wedding:** You have your wedding tomorrow and it is an outdoor wedding. You need to decide whether you should make changes in the organization (like arranging a tent, or moving the wedding indoors).

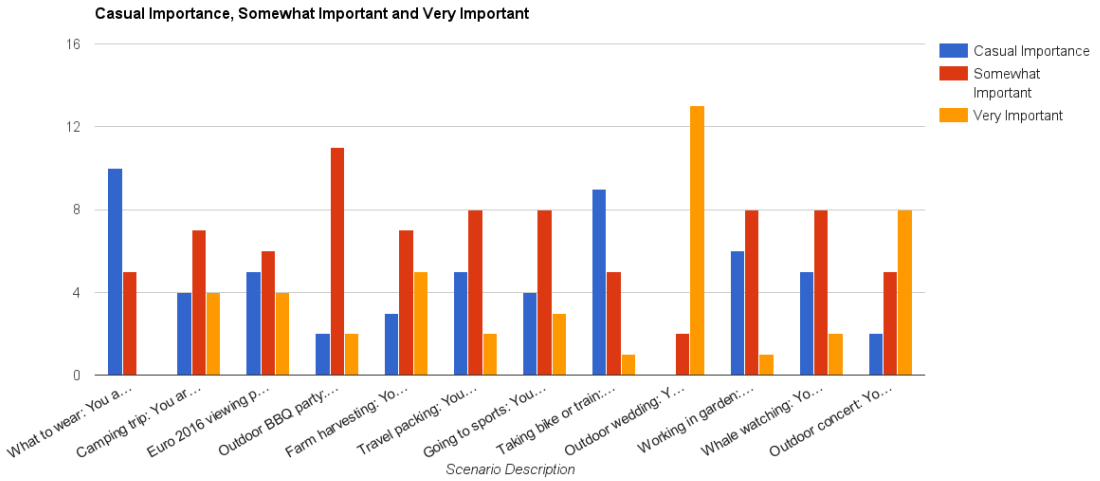


Figure 4.1: Results of the mini survey about scenarios

Time of an event is also another factor while deciding what kind of weather data the user would prefer. People know that the accuracy of weather forecasts decrease and different sources vary more often when the forecast is done for a future date, which might affect the choice of aggregation method. Thus, in order to eliminate the time factor from our scenarios, we have fixed the time for all scenarios to be "tomorrow".

4.3 Questionnaire

4.3.1 Demographics and General Information

In order to make sure that we have covered a wide variety of participants in the survey and to check whether a certain demographic group provides different results, we have asked the following demographic information to each participant:

- Age
- Gender
- Highest level of education

Along with this we have also asked questions about the participant's use of weather forecasts, such as:

- How frequently do the users check weather forecast

- How many weather forecast sources do they use
- Which weather forecast sources do they use
- Why do they use these sources
- Do they regularly compare these sources
- How much do they trust in weather forecasts

This information helped us understand how people use weather forecasts and to see how often do they take advantage of comparing multiple sources.

4.3.2 Evaluation of Aggregation and Representation Methods

We have divided this part of the survey into four sections, for each of the representation methods we have. Our main aim was to compare the aggregation methods with each other to see which one works better in combining multiple weather forecast sources, thus, for each section, we have kept the representation method the same and asked the participants to evaluate the aggregation methods.

At the beginning of each section we have explained the representation method, the scenarios and the aggregation methods. We then showed the participant all the sketches that he/she will be evaluating on this section, so that he/she can compare even the first aggregation method with the others.

Then for each aggregation-representation method combination, we have showed them the sketch we have prepared and asked them to indicate their level of agreement with the following statements, using a 7-point Likert scale:

- I feel confident that this will be the tomorrow's weather
- I would like to use this representation for the scenario: Daily Dress Choice
- I would like to use this representation for the scenario: Outdoor BBQ Party
- I would like to use this representation for the scenario: Outdoor Wedding

After that we have also provided an optional open ended question for the participants to explain the reasoning behind their choice, in order to see why they would pick one aggregation method for a certain scenario.

To make sure that the order we display representation method sections does not affect the answers of the participants, we have randomized the order each participant sees the representation method sections.

4.3.3 Overall Preference

At the end of the survey, we have asked each participant which aggregation-representation method combination they would like to have in their weather application, if they would have to pick just one and we asked the reason behind their choice. Using this information we were able to determine which aggregation and representation method was preferred the most overall. And since this question comes after the participant sees and evaluates all the combinations we have, it is easier for the participant to actually compare them all with each other to pick their favorite. Also by looking at the reasoning behind their decisions, we were able to realize the weak and strong parts of both aggregation and representation methods, which we were able to improve upon for the user study later.

4.4 Participants

Total of 74 participants completely finished the survey whereas other 101 participants left the survey incomplete. For the rest of the incomplete answers we believe that the length of the survey was too much for some people. Another possible reason for the incomplete answers is even though the open ended questions were all optional, most of the participants did not realize that and it could be that the survey can get overwhelming when you answer too many open ended questions. We have also checked to see if there was a specific question or section that people dropped out but there is no particular point in the survey where most of the people dropped out.

Among 74 participants, who have completed the survey, 37 participants were male, 36 participants were female and 1 participant preferred not to state their gender. The average age of the participants was 24.73 with standard deviation of 6.84. The highest level of education for 21.62% master's degree, for 39.19% of the participants was a bachelor's degree, for 25.68% high school diploma, for 1.35% vocational training diploma, for 10.81% secondary school diploma and the final 1.35% of the participants did not have any sort of school diploma. An overview of education levels of participants are visualized in Figure 4.2

9 out of 74 participants stated that they check weather forecast multiple times a day, 23 participants stated that they check daily and 20 participants stated they check multiple times a week. 52.7% of the participants reported they use more than one weather forecast source regularly. Among the participants that use more than one weather forecast source, 58% stated that they are regularly comparing these sources to get a better understanding of the weather and to decide which one is more accurate. The rest of the participants stated that they change the source they use according to the

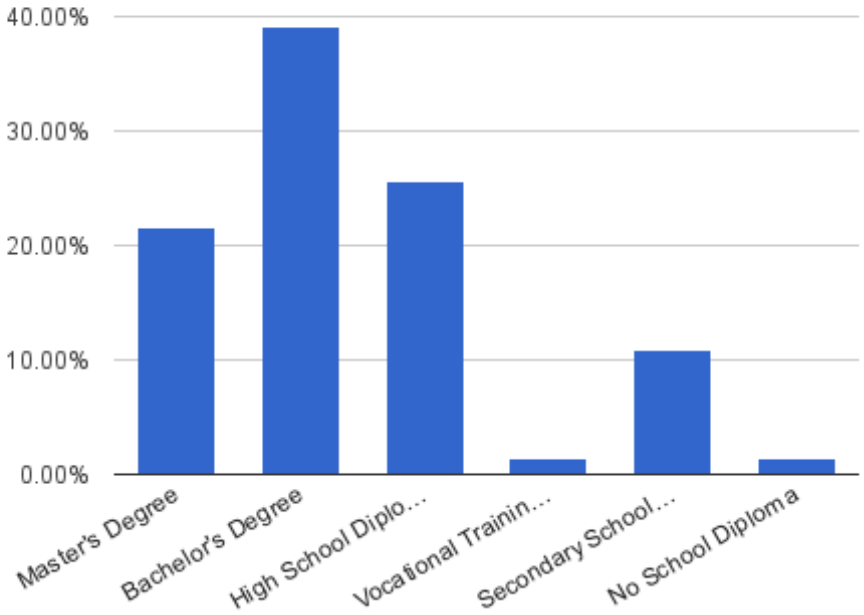


Figure 4.2: Highest level of education of participants represented as percentages

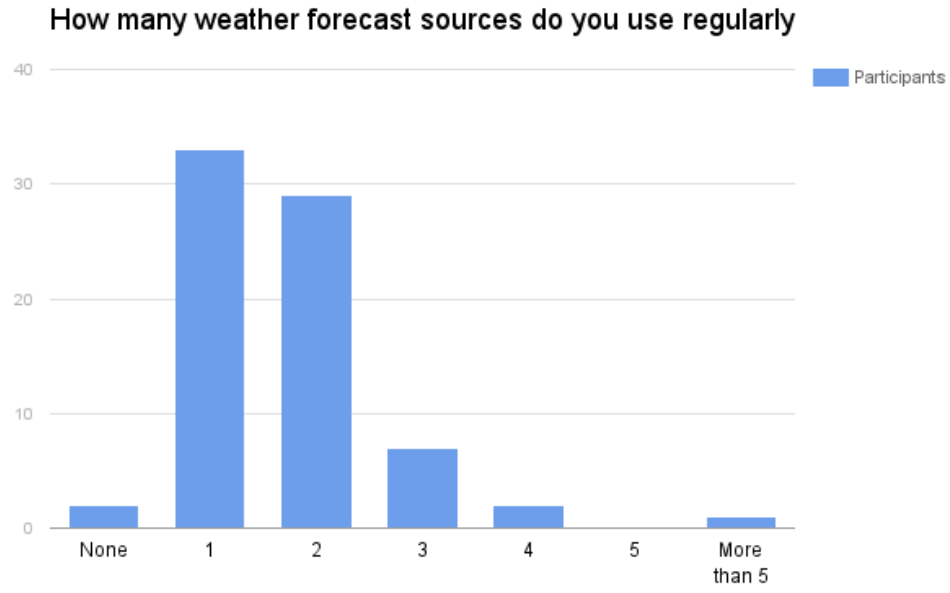


Figure 4.3: Graph of number of weather sources used by participants

device they have closest, if they have a computer they use a website, if they have their smartphone they prefer mobile applications.

4.5 Results

4.5.1 Quantitative Analysis

To begin the quantitative analysis of the results of the questionnaire, we first aligned and rank transformed our Likert item data [WFGH11]. Then using this data we performed two linear mixed-effects model analyses, one for participants' preference ratings and one for participants' confidence ratings. For the analyses for participants' preference ratings, we added the fixed factors aggregation method, visualization and scenario together with the random factor of participant. In the analyses for participants' confidence ratings we added the fixed factors aggregation method and visualization together with the random factor of participant. Significant effects were further explored using post hoc pair-wise comparisons with Bonferroni corrections. Only significant effects of interest are reported.

Aggregation method

Figure 4.4 shows the overall preference ratings of participants for each of the four aggregation methods. The arithmetic mean of the ratings are shown with a vertical red line in each graph. Range has the highest mean rating (M = 4.85, SD = 1.73), followed by direct comparison (M = 4.51, SD = 1.9), mean (M = 4.22, SD = 1.82), and single source (M = 4.13, SD = 1.72). The mixed-effects analysis revealed a main effect of aggregation method on participants' preference ratings (F(3, 3290) = 37.78, $p < 0.001$). Pairwise comparisons of the least-squares means of the linear model showed that participants preferred *range* over the other three methods (range vs. direct: t(3290) = 5.55, $p < 0.001$; range vs. mean: t(3290) = 7.69, $p < 0.001$; range vs. single: t(3290) = 10.21, $p < 0.001$), as well as *direct* over *single* (direct vs. single: t(3290) = 4.66, $p < 0.001$).

Figure 4.5 shows the mean confidence ratings of participants for each of the four aggregation methods as a function of aggregation method. Range has the highest mean rating (M = 5.34, SD = 1.46), followed by direct comparison (M = 4.9, SD = 1.75), mean (M = 4.83, SD = 1.46) and single source (M = 4.79, SD = 1.5). The linear mixed effects analysis revealed a main effect of aggregation method on participants' confidence ratings (F(3, 1050) = 13.5, $p < 0.001$). In particular, pairwise comparisons showed that

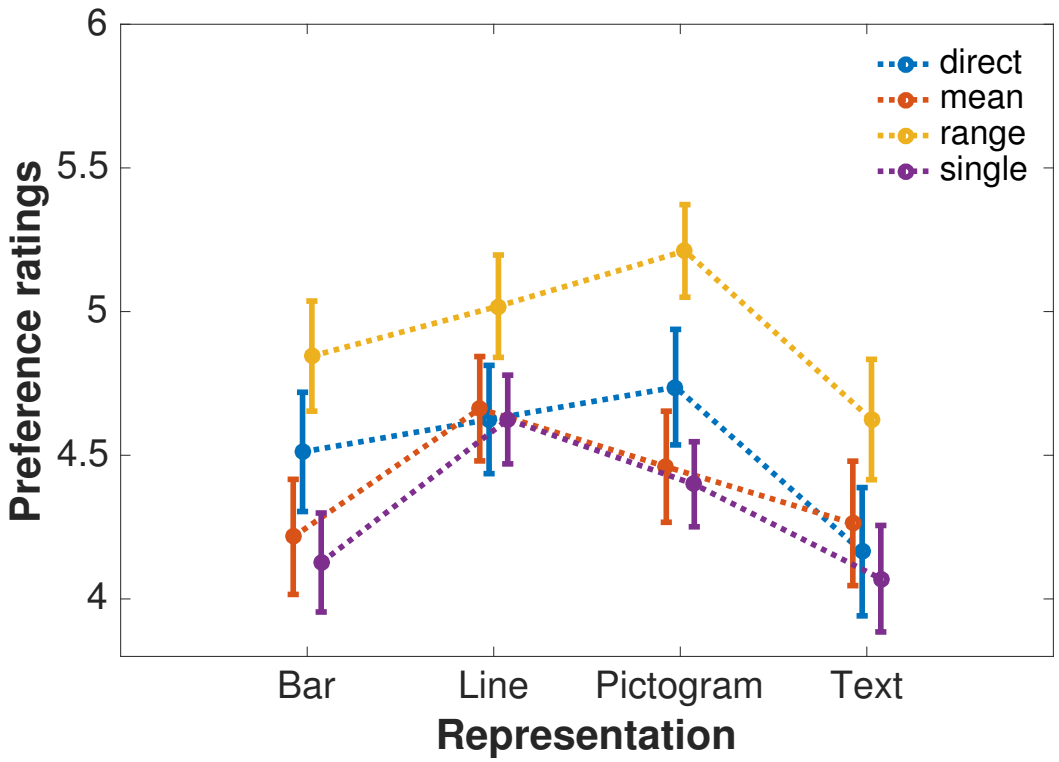


Figure 4.4: Overall preference ratings of participants for each aggregation method

participants indicated higher confidence in *range* compared to the other three methods (range vs. direct: t(1050) = 3.48, $p < 0.01$; range vs. mean: t(1050) = 4.85, $p < 0.001$; range vs. single: t(1050) = 6.0, $p < 0.001$).

The results from the quantitative analyses states that people generally prefer to receive more information rather than less information in the context of weather, in other words, people prefer aggregated presentations over single source presentations.

Representation

Figure 4.6 shows overall preference ratings of participants for each of the four representations. The mean rating is marked with a red line on each graph. Line has the highest mean rating (M =4.73, SD = 1.67) followed by pictogram (M = 4.7, SD = 1.73), bar graph (M = 4.42, SD=1.81), text (M = 4.28, SD = 1.88). Using the mixed-effects analysis we have revealed a main effect of representation on participants' preference rankings (F(3, 3290) = 16.48, $p < 0.001$). Pair-wise comparisons of the least-squares means of the linear model indicate that participants preferred *line* and *pictogram* over *bar* and *text* (line vs. bar: t(3290) = 4.17, $p < 0.001$; line vs. text: t(3290) = 5.92,

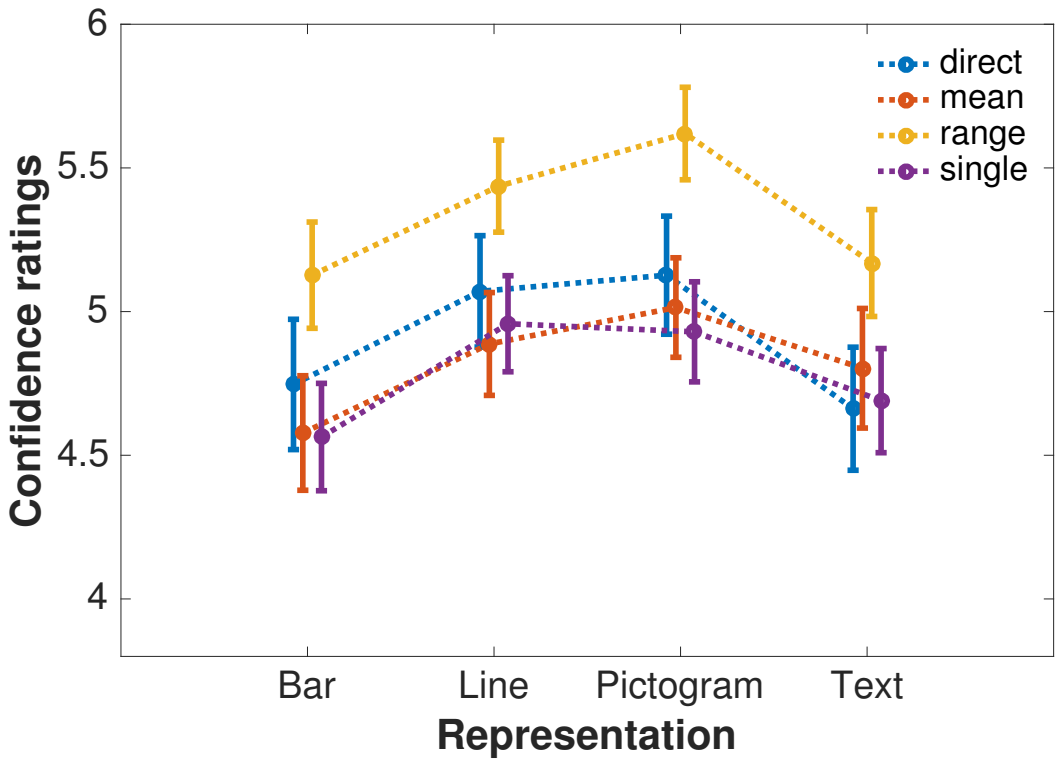


Figure 4.5: Overall confidence ratings of participants for each aggregation method

$p < 0.001$; pictogram vs. bar: t(3290) = 3.682, $p < 0.001$; pictogram vs. text: t(3290) = 5.436, $p < 0.001$).

Figure 4.7 shows overall confidence ratings of participants for each of the four representations. The mean rating is marked with a red line on each graph. Confidence ratings of mean were highest for *pictogram* (M = 5.17, SD = 1.53), followed by *line* (M = 5.09, SD = 1.49), *text* (M = 4.83, SD = 1.67), and *bar* (M = 4.7, SD = 1.69). The linear mixed effects analysis showed a main effect of representation (F(3, 1050) = 7.27 , $p < 0.001$). Pair-wise comparisons showed that *pictogram* received higher confidence ratings than *bar* (t(1050) = 4.08, $p < 0.001$) and *text* (t(1050) = 3.8, $p < 0.001$). By looking at these results we can conclude that the participants preferred the representations *line* and pictogram. The *pictogram* representation also receives a confidence bonus.

Interactions between aggregation method, representation, and scenario

Among aggregation methods, representation methods and scenarios, there are three possible two-way interactions and one three-way interaction. Among these interactions between the three factors in the three-way analysis of **preference ratings**, only the

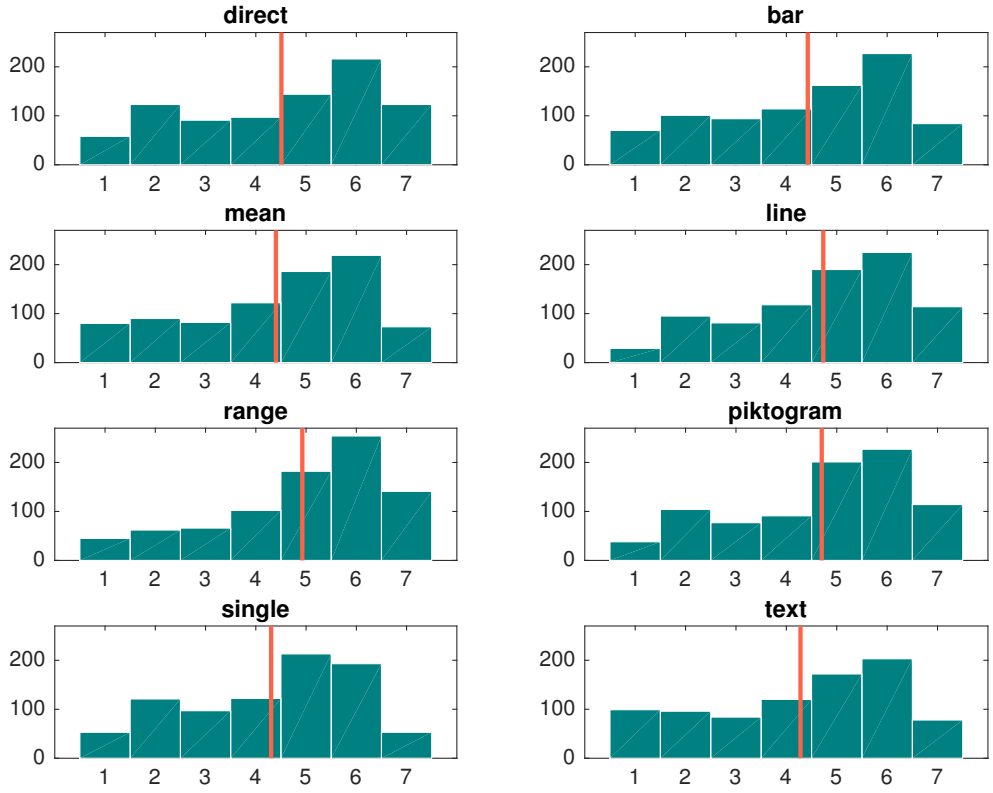


Figure 4.6: Overall preference ratings of participants for each representation method

interaction between *aggregation method* and *scenario* is significant (F(6,3290) = 11.71, $p < 0.001$). The mean preferences for each of the four aggregation methods as a function of the scenario are shown in Figure 1.d. The data of each scenario was centered on the scenario's overall mean, for illustration purposes. The reason for this was since preference differences between the three scenarios are of no interest to us (i.e., the main effect of scenario) and may mask trends in preference ratings across the three scenarios.

The main source of the interaction effect is the increase in the preference ratings for the *direct* aggregation method as the importance of the scenario increases, which is also visible in the figure. Out of the total nine pairwise comparisons, the five that involves *direct* aggregation method show significant differences of the difference of participants' preference ratings between the *direct* aggregation method and one of the other methods across two levels of the factor scenario (e.g., for the daily scenario *single* presentation is preferred over the *direct* aggregation while the situation is reversed for the wedding scenario; see Table 1 for test statistics and p-values for all significant

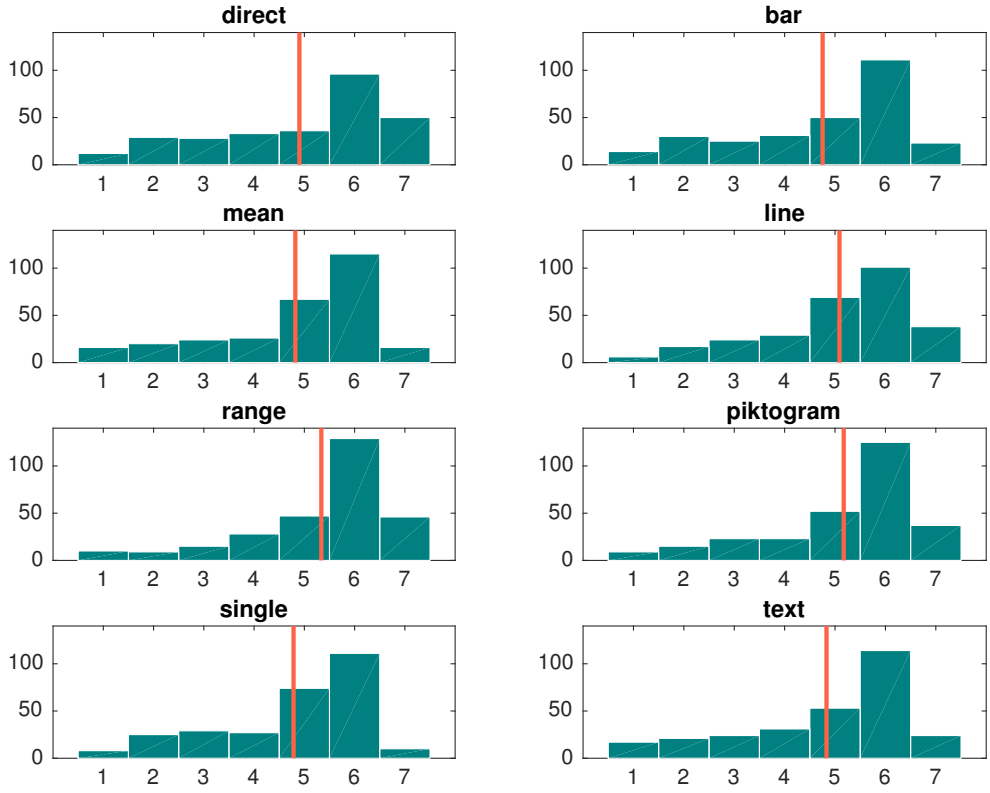


Figure 4.7: Overall confidence ratings of participants for each representation method

contrasts). This partially means that the user's willingness to spend time and effort to perform aggregation manually, increases with the importance of the scenario that the weather forecast is used in (yet, the range aggregation method still remains as the overall preferred method).

A secondary source of interaction was found that results from the stronger decrease in preference of the *single* source as compared to the other aggregation methods as importance of scenario increases (e.g. the difference between *single* and *range* is larger for the wedding scenario then for the daily scenario; see Table 1 for test statistics and p-values for all significant contrasts). This confirms that the preference for more information (i.e. aggregation of several weather sources) increases with the importance of the scenario.

Contrast	df	χ^2	p-value
Direct-Single : BBQ-Daily	1	22.47	< 0.001
Range-Single : BBQ-Daily	1	10.35	0.023
Direct-Single : BBQ-Wedding	1	11.37	0.013
Direct-Mean : Daily-Wedding	1	25.24	< 0.001
Direct-Range : Daily-Wedding	1	10.23	0.025
Direct-Single : Daily-Wedding	1	65.81	< 0.001
Mean-Single : Daily-Wedding	1	9.54	0.036
Range-Single : Daily-Wedding	1	24.15	< 0.001

Table 4.1: Difference of the differences between two aggregation methods across two scenarios (e.g., is the difference between Direct and Single larger in the BBQ or in the Daily scenario).

4.6 Discussion

The results from the survey generally shows that the participants preferred to have aggregated data rather than data from a single source. Among the aggregation methods, mean had the smallest preference rating score. Participants preferred the aggregation methods range and direct comparison since they provide more transparency on the values from individual sources, and thus allows making more informed decisions. Some participants stated their concerns about the mean aggregation methods, telling it would be difficult to spot an extreme value among the sources that might affect the calculated mean, which makes it less reliable. Besides mean aggregation method does not provide any additional data then single source, the only difference is the user knows the data comes from multiple sources, yet the results show that this is not enough motivation for users to feel more confident about it.

As for the representations methods, line graph and pictogram were the most preferred ones. We believe pictogram were picked because of people's familiarity with using pictograms in weather forecasts. Based on some comments written by the participants, we conclude that line representation was mostly preferred because that was the only representation that provided hourly information, and for the participants that find hourly weather information very valuable, the choice was based on the data they get rather than the representation itself. Thus, we have decided that pictogram representation improved with hourly weather information would be the ideal representation method.

We did not find any interactions between the aggregation and representation methods. This allowed us to make our decisions based on the overall preference of the participants, since the preference of aggregation methods were completely independent of the

representation method used. This also cleared our concerns about possible bias, the participants would have, due to possible synergy between an aggregation method and a representation method.

We have found an interaction effect between the aggregation method and the scenario, which shows us that as the importance of the scenario increases, the preference for range aggregation method slightly increases and the preference for direct comparison aggregation method greatly increases, while the preference for single source drops below all other aggregation methods. This shows that people prefer having more information and more control over the data when making important decisions, and they prefer aggregated methods over a single source in this case. And also it shows that the range aggregation method provides enough details for the users to feel confident about even their important decisions.

5 Android Application

After our survey to determine which kind of aggregation and representation works better, we have built an Android application to test our findings in real life conditions.

5.1 Requirements

Based on the results of the survey, we have decided to implement the application to use the range aggregation method in combination with pictogram visualization. Looking at the comments and suggestions from the survey, we came to realize that, some participants said they prefer the line graph visualization for the sole reason it contained hourly information. Thus, we decided that the application should contain hourly information but it should still use pictograms for visualizing the hourly data. In addition since people check the weather forecast mostly a day or two days in advance, we have decided that the application should provide hourly three day weather forecast.

The survey results showed that participants preferred the direct comparison aggregation method for especially important events. Therefore, we have decided to combine direct comparison aggregation method with range aggregation method so that the users can toggle between the two aggregation methods on demand. Range aggregation method would be the default view where the user can decide to see more detailed information by tapping on a button to switch to the direct comparison aggregation method. In order to be consistent with the changes to the range aggregation method we have also included the hourly data in the direct comparison aggregation method.

5.2 Design

For the final design, we have built upon the layout of our existing sketches used in the survey. With the addition of hourly data to the existing layout, we needed to find a solution for both the range and the direct comparison aggregation methods. In addition to this, we needed to make sure the application would look good on variety of different devices and screen sizes. And finally we needed to update the look of the application

Figure 5.1: User can switch between today, tomorrow and the next day using the tabs on the top of the screen.

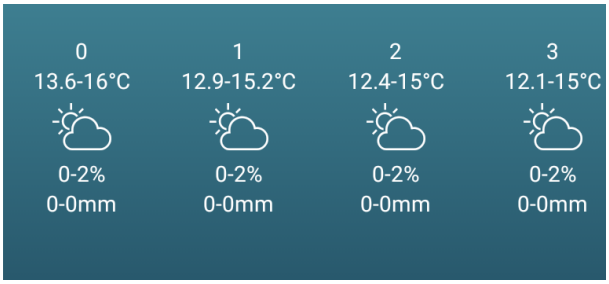


Figure 5.2: Hourly data section in range aggregation method. The view is horizontally scrollable so that the user can browse the hours of the day.

so that it feels like a professional application to the user study participants, rather than a simple prototype; which is important for the participants to easily replace their old weather application with ours.

5.2.1 Hourly Data

Presenting hourly data for three day weather forecast on a small screen is not feasible if the design is static. Thus, we have used interactions to make it easier for the users easily consume this information. First we have separated each day into its own tab, and provided a tab navigation at the top of the screen to switch between the days as seen on Figure 5.1. For the range aggregation method we have used a horizontally scrollable view at the bottom of the screen which shows the hourly data for the selected day and the hourly data contains the temperature, chance of rain, amount of rain and the pictogram for that specific hour. The data starts on the current hour, labeled as 'now' and by scrolling horizontally, the user can see the weather for each hour of the day. Figure 5.2 shows the final design of the hourly data for range aggregation.

Figure 5.3: Hourly data section in direct comparison aggregation method. The hour changes when the user swaps right or left, at the same time for each user so it is comparable.

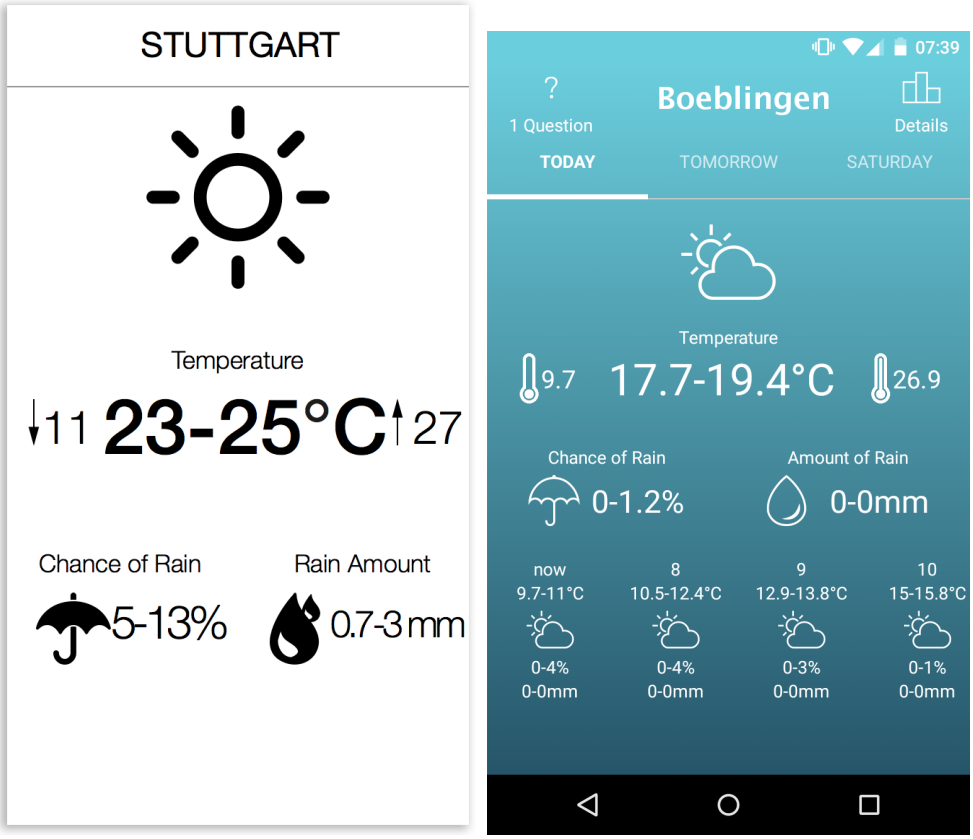
For the direct comparison aggregation method, we wanted to make sure that the user can easily compare the hourly data from sources with each other. Thus, we have provided the data for one hour below each source and by swiping left and right, the user can switch between the hours of the day. This way the hourly data from each source is always in sync for users to compare them. The final design of the hourly data for direct comparison aggregation method is shown in Figure 5.3

5.2.2 Updating the Look

The purpose of the sketches we have prepared for the survey was only to test the concepts and to see which aggregation and representation method was better in combining multiple weather forecast sources in one application. Hence, the sketches were prepared as low fidelity mockups in order to make sure that the participants base their decision only on the used aggregation and representation methods rather than their personal liking of one of the visuals.

For the user study, we wanted participants to replace their existing weather application with our application. Thus, we wanted the application to feel like a professional weather application rather than a study prototype.

To begin with, we have changed the background color to a gradient of two shades of blue and set the font color to white for better readability. In order to make sure the screen was easily readable on various screen brightness amounts and environments, we ran some quick tests with three different devices and adjusted our colors to be visible in these conditions. We have used Lucida Grande as the typeface and adjusted the font sizes to make sure everything looks properly on different screen sizes. The final design of the screen is shown in Figure 5.4 and 5.5 in side to side comparison with the original sketch used in the online survey for both range and direct comparison aggregation methods.



(a) Original Sketch **(b)** Final Design

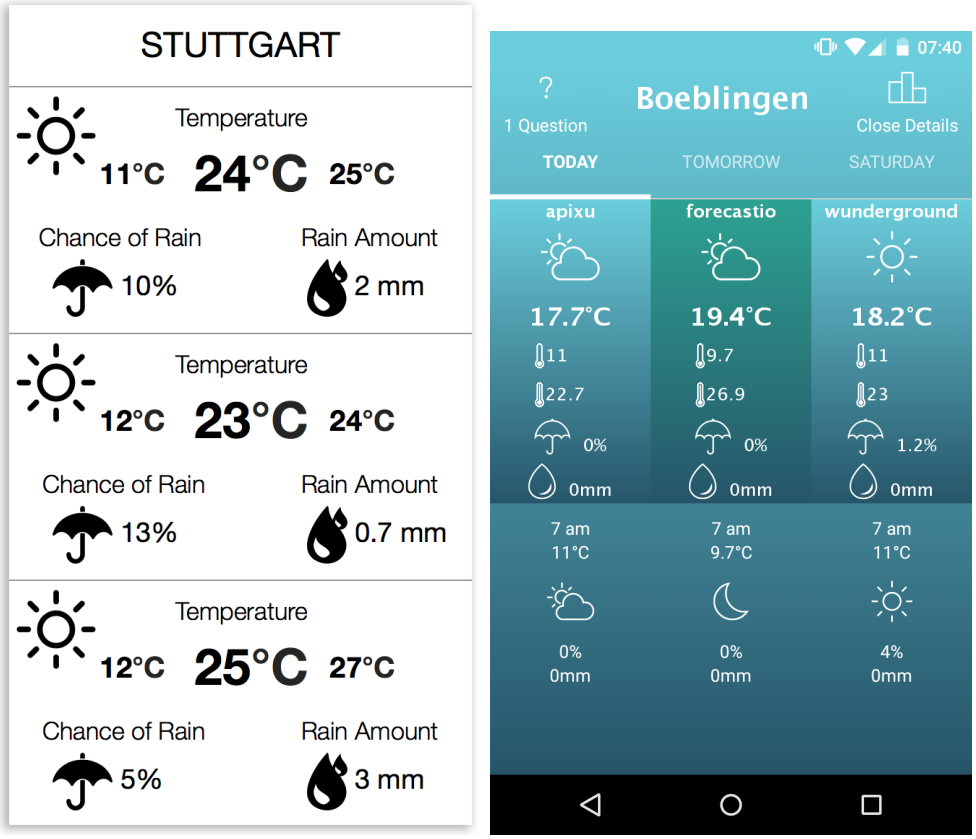
Figure 5.4: The original sketch used in the online survey and the updated final design for range aggregation with pictogram representation

5.3 Development

5.3.1 Technology Stack

The Android client is built using React Native[1], which allows creating native mobile applications using Javascript and React[2]. The reason we have preferred React Native over the native Android development environment is the speed of development that comes with React Native, while still providing the performance and feeling of native Android.

[1] **React Native:** https://facebook.github.io/react-native/
[2] **React:** https://facebook.github.io/react/



(a) Original Sketch **(b)** Final Design

Figure 5.5: The original sketch used in the online survey and the updated final design for direct comparison aggregation with pictogram representation

The application is backed up by a server which retrieves data from third party weather APIs and also logs the necessary information to a database for the purpose of the user study. The back-end server is written using Node.js[3] using the Express.js web framework [4]. It uses MongoDB[5] for persistent data storage. The back-end server provides a RESTful API which is used by the Android client and returns JSON responses for each request over HTTP.

[3]**Node.js:** https://nodejs.org/en/
[4]**Express.js:** https://expressjs.com/
[5]**MongoDB:** https://www.mongodb.com/

5.3.2 Android Client

The Android client is the user facing part of the application. It is responsible for rendering the user interface and reacting to the user input. When the user opens the application for the first time, the client stores a participant unique UUID which is used to identify the client in the server logs. Along with the UUID it also stores the date the user started using the application.

The client gets the location of the user on startup and communicates with the back-end server REST API asynchronously. When the server returns the data from weather providers, the client aggregates the data to display according to the active view.

For the purpose of the study we log certain events that happen while the user is using the application. The client calls the back-end server to record whenever the user opens the app or switches between the two aggregation methods. Also when the user is leaving the application, or if he/she taps on the question mark icon on the top left part of the screen, the user is prompt with one question daily, which is needed for the study.

The architecture of the Android client follows the Flux pattern[6]. User interactions trigger actions, actions are responsible for creating and dispatching events which contain the information about what needs to be done in response to this interaction and the data needed to do it. When an event is dispatched by an action, the dispatcher, which follows a modified version of the publish-subscribe pattern, sends this action to all the stores in the application. In our application we only have one store called Weather Store. Weather Store is responsible for keeping the latest weather data. When the application is first started, it contains temporary data. Then whenever a new event is received, the store updates the data it keeps, according to this event and then emits a new event to notify the view that the store has changed. For example when the application is loading at the beginning, the corresponding action dispatches an event with the weather data retrieved from the server. Upon receiving this event, the store overwrites its temporary data with the data from the server and emits the event telling the view that the store has updated. The view is always listening the store for any change and whenever it receives the event, it gets the data from the store and updates the necessary parts in the user interface. React Native does not completely render the whole screen every time the data changes, instead it just updates the parts of the user interface that were affected by this change in the data. Rather than retrieving the raw data from the store, we have returned a service that provides functions which return just the necessary parts of the data that the user interface needs. This way we were able to keep our views clean and do the data manipulations and aggregations inside this service.

[6] **Flux pattern:** https://facebook.github.io/flux/

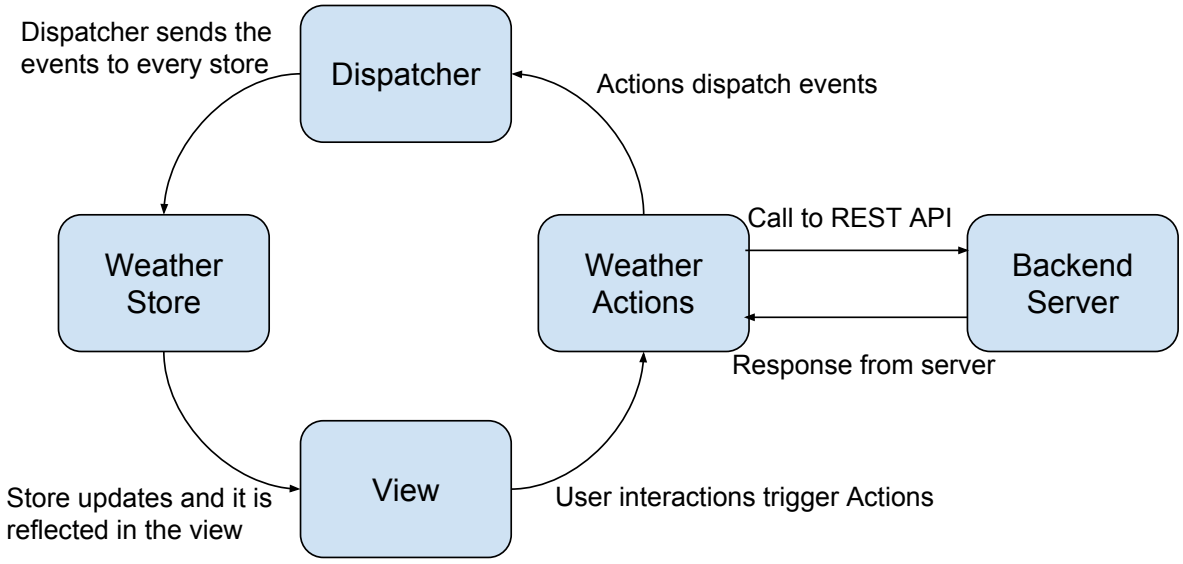


Figure 5.6: Flux architecture used in the Android client

5.3.3 Back-end Server

The back-end server is responsible for providing the weather data for the Android client and logging the user events for the user study. The server provides the following functionality over a RESTful API:

- Retrieve weather data from three different sources
- Retrieve weather data from a single source
- Post, retrieve, update and delete questions for the user study
- Post answers to questions
- Create a log record for user opening the application
- Create a log record for user switching between the aggregation methods

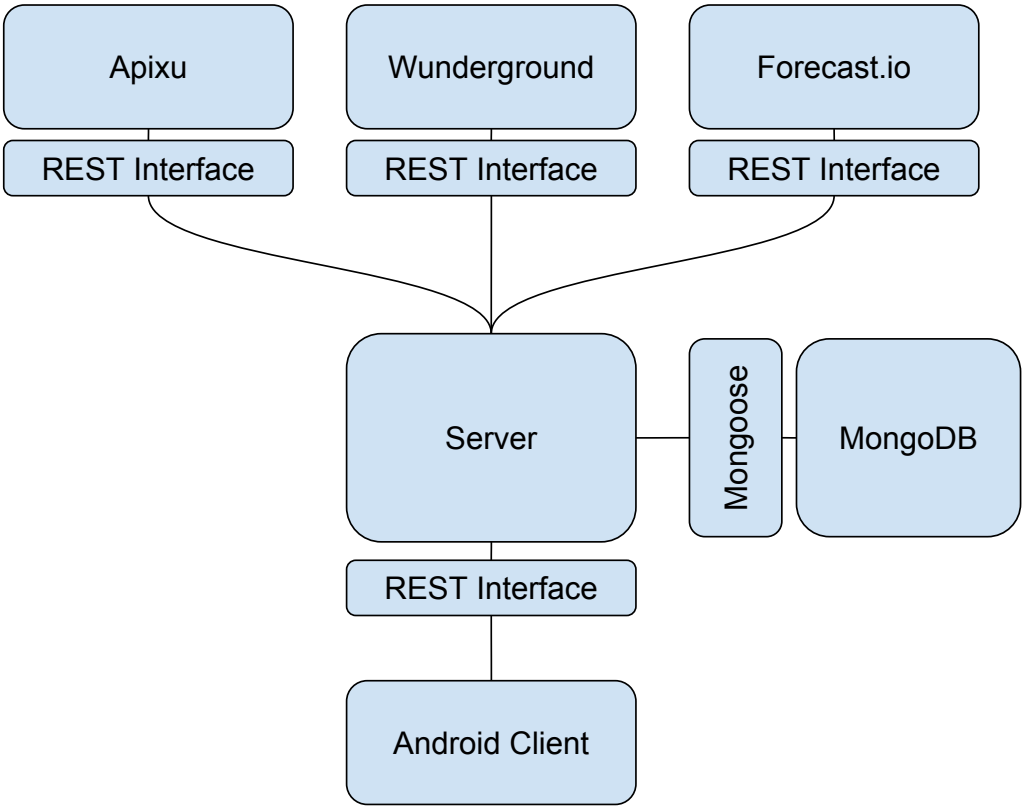


Figure 5.7: Overall view of the server structure

The server uses REST APIs provided three separate weather forecast providers, Forecast.io[7], Wunderground[8] and Apixu[9]. When the client makes a request to retrieve the weather data by providing the latitude and longitude of the location, the server uses a geocoder service to locate the closest city and makes a request to all three weather forecast providers. These calls are made asynchronously so each call happens simultaneously, which shortens the response time of the server. Upon receiving the raw data from these three sources, the server maps the information into its own model for weather data in order to standardize the information and make the server decoupled from the weather forecast sources. This way the server can easily handle changes, additions and removals of weather forecast sources.

[7]**Forecast.io:** https://darksky.net/dev/
[8]**Wunderground:** https://www.wunderground.com/weather/api
[9]**Apixu:** http://www.apixu.com/

Table 5.1: Summary of selected Weather Forecast Providers

Weather Provider	Hours of forecast	Required fields	# of free API calls
Apixu	240	+	5000 per month
Forecast.io	48	+	1000 per day
Wunderground	240	+	500 per day

5.4 Weather Sources

In order to test our aggregation and representation methods properly during our study, we needed to use real weather data. By relying on third party sources to provide the weather data, we made sure that the study is only measuring the effectiveness of our aggregation and representation methods rather than measuring the accuracy of the weather data, since the third party sources we have used are all reliable and trusted sources that already have a lot of users.

By looking at the requirements of our Android application that we have gathered from the results of the survey, we derived the following set of requirements that a third party weather forecast API should fulfill in order to be integrated in our application:

- Should provide hourly weather data for at least 3 days (72 hours)
- Should provide temperature, chance of rain, amount of rain and a pictogram
- Should allow reasonable amount of API calls per day
- Should be free to use for non-commercial/research purposes

After considering several weather forecast APIs, we have found Forecast.io, Apixu and Wunderground APIs to be the APIs that fits most with our requirements. However, the API provided by Forecast.io only provides 48 hour weather forecast but since we could not find any other API that satisfied our requirements better, we have decided to keep the API from Forecast.io and instead change our application to work with 48 hours only. In order to keep our sources consistent, we have also trimmed the output from other two APIs to 48 hours. Even though this is not the ideal case for the application, this was the best approach for us to run the study within the given time and budget. And since our survey was also considering only the scenarios for the next day, 48 hours of weather data was sufficient for our study. The comparison of the three weather forecast APIs can be seen in Table 5.1.

6 User Study

After the initial survey to evaluate the aggregation and representation methods, we have came up with certain requirements and implemented an Android application that satisfies those requirements. The purpose of the user study was to see how selected aggregation and representation methods developed in this Android application work in real life scenarios, when people use it daily.

6.1 Study Composition

We wanted to make sure that we get feedback and usage statistics from people who are using our application for their weather forecast needs. Thus, we have asked each participant to replace their existing weather forecast application with our application throughout the duration of the study which was one week. During this time we asked them to use the application as usual and answer a question that asks the user's agreement on the statement "I feel confident about how the weather will be" using a 5-point Likert scale once everyday after they use the application. The question would automatically pop up when they try to leave the application.

During the course of the study, participants used the application for a week, and the application would show the weather using three different sources combined with range aggregation method and pictogram representation method, together with the ability to toggle details view which provides the individual values from each source using the direct comparison aggregation method.

After one week, we have provided the participants with the survey we have prepared to get feedback on the application they had been using. The details about the survey is discussed in the next section.

6.2 Tasks and Questionnaire

Since we wanted to measure the effectiveness of our aggregation and representation methods in natural, real life circumstances, we did not provide certain tasks to the participants. Instead we had our participants using the implementation of our methods as an Android application, integrated into their daily life. We specifically had the requirement for participants to be in their usual routine during the study and we did not allow anyone to participate if they were planning to take a vacation since that would change the behaviour of the participants both because they might care more about the weather when in holiday and also it could be that they do not have internet connection during their holiday which would mean they would not be able to use the application and we would not be able to gather any usage statistics.

Alongside checking the weather regularly, participants were required to answer a question to give feedback on how they feel about using the application. The questions were provided by the application at the moment the user attempts to leave the application. The users were also notified by a question mark icon at the top left corner of the application when they were required to answer a question, which they could immediately answer by tapping on the question mark icon. The questions were provided daily for each participant and whenever the participant answers a question for that day, they would not see the question the next time they use the application within the same day. Figure 6.1 shows the question dialog and how to open it.

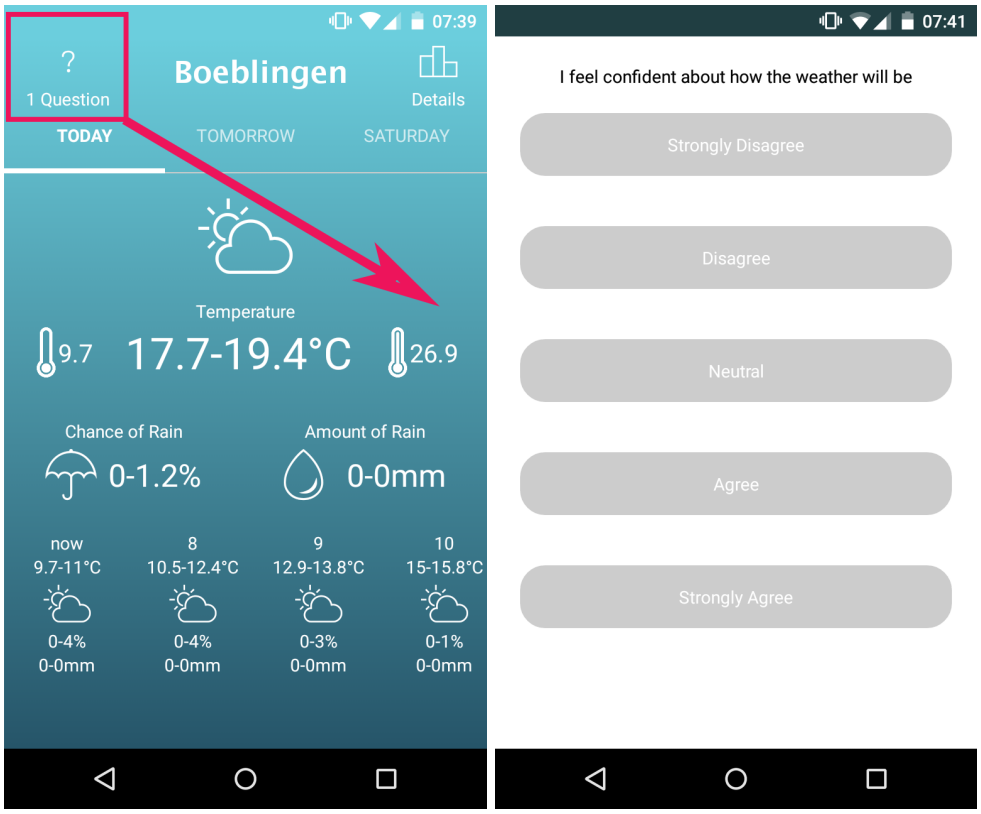
After one week of using the application, we have presented the participants with a questionnaire about the application they had been using.

6.2.1 The Questionnaire

For the participants we have prepared a questionnaire to get some feedback about what they think about the application. The questionnaire consists of three sections, demographic data, standardized feedback and general feedback.

Demographic data

In demographics data section, we gather basic information about the participant such as the age and gender, which let us understand whether we have covered a wide and evenly distributed variety of participants.



(a) The button to open the question **(b)** The view where the user answers the question

Figure 6.1: The daily question that the app provides to each user

Standardized feedback

Since we did not want the survey to take a lot of time for participants and since we did not have huge number of participants (only 21), we have decided to use System Usability Scale (SUS) [Bro96] in our questionnaire. SUS enabled us to get a measure of the perceived usability of our application with a small sample of users and in a short time. SUS contains 10 statements where the participant is required to indicate their level of agreement using a 5 point Likert item. The statements are aimed at measuring the participants' ideas about ease of use, learnability and complexity of the application. The questions in this section were mandatory to answer and since we have used a standardized questionnaire, we were able to compare the results of our application with other systems in terms of usability for each participant.

General feedback

The general feedback section aims to collect the opinions of participants regarding the application using open ended questions. This allows us to understand the thought process of the participants, what they liked, what they did not like, what do they think about having multiple weather forecast sources in one application, what would they want to change or add to the application to make it better and whether they would actually want to use it in the future. By providing this section, we wanted the participants to share their honest opinions which can help us improve the application and might help us discover future research focuses related to this topic.

6.3 Participants

A total of 21 participants were involved in the user study. Among these 21 participants, 43% were female, 52% were male and 5% did not want to provide information about their gender. The average age of participants were 30.75 with the standard deviation of 10.78. One of the participants decided not to provide their age.

All the participants had Android phones with varying versions and they all stated that they check weather forecast regularly in their daily lives, they are not and will not be on a vacation while the study is in progress.

In order to make sure that the participants use the application exactly for one week, we made the tracking of the usage on our server and the application stopped working for them after one week. By doing this we eliminated the possible issues where a participant might take a few days to install the application after we send it to them over email. We have kept the final questionnaire active for a week so that every participant were able to fill it after they have used the application for exactly one week.

6.4 Results

The results of the user study will be evaluated in three sections, the results from System Usability Scale (SUS), the comments from the general feedback section of the survey and the usage statistics from the Android application logs.

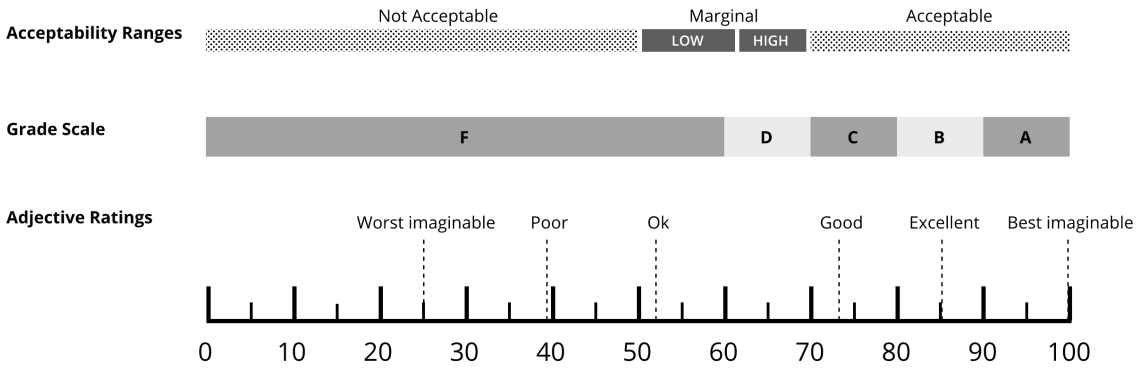


Figure 6.2: The SUS rating scale by Bangor, Kortum, and Miller [BKM09]

6.4.1 Results of System Usability Scale

The analysis of System Usability Scale (SUS) results in a value between 0 and 100, which indicates the perceived usability of a product by the users. SUS has been used by many researchers and companies over the years which allows us to compare our results with theirs. Bangor, Kortum, and Miller [BKM09] have compared over 3,500 SUS results and came up with a rating scale which maps SUS scores to adjective ratings, grade scales and acceptability ranges. The grading system can be seen on Figure 6.2.

The average score of our SUS questionnaire among 21 participants were calculated as 70.4 which, according to the scale, corresponds to “Good” in adjective ratings, “C” in grade scale and “Acceptable” in acceptability ranges. The score shows that participants perceived the usability of our Android application as *Good* in general, yet there is definitely room for improvement in that term.

6.4.2 Findings from General Feedback

The general feedback section of the questionnaire allowed participants to provide their feedback in open text format. This way we were able to get suggestions and learn about the weak points. Overall we can group the findings from the general feedback section of the questionnaire into 3 categories, suggestions, positive feedback and negative feedback.

Suggestions

One of the things that most users wanted was to have the ability to get the weather forecast for multiple locations. This is a very common feature for a weather forecast application and the only reason we did not include it in our application was to keep the development time short. Along with this feature, participants also requested that the application should work without using GPS to save battery. So in general the application should keep a list of cities for the user and only when requested, access GPS data to show weather for the current location.

The second most requested feature was to have more than 3 days of weather forecast. Most of the participants stated that they would like to have the weather forecast for at least a week. The reason we have limited the application to only 3 day weather forecast was the limitations forced by the weather APIs we are using. Ideally the prefered way would be to provide the weather forecast for at least one week together with the hourly data for each day.

The third most requested feature was to provide a widget for quick access from Android home screen. Most of the participants stated that they are using a widget for their weather application, and only open the application itself when they want more detailed information. We have not considered creating a widget for the user study but as requested by the users, but acknowledge that having a widget that only shows the overview of the weather forecast using the range aggregation method would be useful for the users.

Positive Feedback

Users generally agreed that having weather information from multiple sources increased their confidence level. Users stated that even though they do not know how accurate the weather information is, they have felt more confident using Weather Total compared to their usual weather forecast application.

Most of the users stated that they liked to use the range aggregation method. They have stated that they used the range aggregation method most of the time and only used the details view with direct comparison aggregation method in certain times. Some claimed they only opened the details view out of curiosity to see what values come from which source, some said they used it only if they were planning to go out, and some said they have only used the details view when the values in the range aggregation was too wide.

There were also some users who claimed that they have only used the details view with direct comparison aggregation method, since they wanted to have more information about the weather all the time.

Users also stated that they liked that the application provides precise hourly information for each day. For most users hourly data was more important than daily average, since they wanted to know the weather for exactly the hours they will be outdoors.

Negative Feedback

The major issue users had with the application was that there is too much information shown on the screen and most of the participants found the screen very crowded. Some argued that they did not want to see the hourly data all the time, others said having only the average or the maximum temperature for the day would be enough rather than showing minimum, average and maximum temperature. Some suggested the rain amount would be omitted since they did not found it really useful. Yet, in general they all agreed that the main screen contains too much information and it should be clearer for better usability.

Another issue that were reported by most of the users were that the horizontal scrolling required to browse through the hourly data was too much. They claimed it took too long to check the weather for tomorrow afternoon. A possible solution to this issue would be to make the representation of weather data for one hour smaller, so that more hours would fit on screen without scrolling, and amount of scrolling needed would be less.

Some users also stated that having weather data with digits after decimal point were making the screen very crowded and they have claimed that it did not increased their confidence in the accuracy of the weather forecast.

6.4.3 Usage Statistics

The Android application keeps track of the basic interactions in the application. It logs the time and the user id for when the user opened the app, switched to details view and switched back to the range aggregation view. By looking at these logs, we can derive some statistics about participants' usage.

During the one week period of the user study, participants have opened the app on average 17.95 times, with standard deviation of 5.72. This statistic includes any time that the application was opened in the participant's phone, thus it is possible that there could be cases where they have opened the application by mistake or opened the app

without enabling location and then had to restart because the location was not found. Yet, we can still say that the users were using the application for more than once a day.

During this time, participants have opened the details view on average 6.14 times, with standard deviation of 2.69. There was a reported issue in the Android application used for the study, where sometimes it took a few seconds to switch to the details view and participants stated that there were times that they tapped the details view icon, and then tapped again because they could not see immediate feedback. Which resulted in details view showing up and then disappear immediately. Thus when the user taps again to actually open the details view, the application would log as if the details view was opened two times. Considering this issue as well, we can say that participants did not open the details view every time they have used the application.

6.5 Discussion

Overall the user study showed us that users like being able to see information from multiple weather forecast sources using only one application, and the range aggregation method combined with direct comparison aggregation method for details view is successful in combining these sources. By looking at the usage statistics and people's comments from general feedback section, it is safe to say that for most of the cases range aggregation method provides enough information for the user, and the direct comparison aggregation method is only needed when the user needs to make an important decision based on weather or if the user has personal interest in weather forecasts and likes to see as much information as possible.

As suggested by the users, the application should be improved to support multiple locations and at least week-long weather forecast, along with some performance improvements and bug fixes. Most of the users stated that they would like to replace their existing weather application for Weather Total given that the mentioned features were implemented

7 Summary and Conclusions

In this section we provide a brief summary of the goals of this thesis, what we have done and what we have achieved. Then we will present our conclusions, followed by suggestions and possible future work.

7.1 Summary

We have started our research by examining the existing literature, investigating about popular weather applications to see what they do right and what they do wrong according to users' feedback, and researching whether there are existing research or commercial applications that utilize our topic of aggregating multiple weather forecasts. After the initial research, we have found that there were only two commercial applications that followed a similar idea of combining multiple sources and there were no academic literature about these applications. After investigating all the sources we have found, we have started preparing sketches for the ideas we had based on the initial research. At the end of researching and sketching possible ideas, we have found 4 aggregation methods and 4 representation methods worthy of investigating more. The aggregation methods are Single Source, Mean, Range and Direct Comparison. We have included Single Source as an aggregation method in our research in order to be able to compare our methods with the existing weather applications that use only one source. As for the representation methods, we have decided to investigate Pictogram, Line graph, Bar graph and Text visualization methods for each of the aggregation methods we have.

With the four aggregation and four representation methods, we had 16 possible combinations which we wanted to test in order to see which one is more preferred by the users. In order to find out, we have conducted an online survey where we have asked participants their preference of aggregation and representation methods with additional factor of scenarios, where we make them consider each sketch for a scenario of varying importance. The results of the survey showed that participants preferred the range aggregation method combined with the pictogram representation method the most. Along with this we have also found out that when the importance of the scenario increases, participants' preference of using direct comparison increases a lot. Thus,

we have decided to continue our research with range aggregation method combined with pictogram representation method and with the option to see details on demand using direct comparison aggregation method. And from looking at the participants' comments in the survey, we have realized that most participants value hourly weather information greatly, thus we have decided to include hourly weather information in our aggregations.

Based on the results of our survey, we have developed an Android application using React Native and Javascript which was supported by a Node.js back-end server. The application aggregated weather data from three third party weather APIs, and presented that to the user in range and direct comparison aggregation methods. Using this Android application, we have conducted a user study, where participants used our application instead of their existing weather application for 1 week and then gave us feedback on what they feel about using aggregated weather information. Based on the feedback we have received as the result of the user study, we have came to the conclusion that people prefer aggregated weather forecasts over a single source weather forecast, having information from multiple sources increases the confidence of the user, people are satisfied by the aggregation provided by range aggregation method in general and most only use the direct comparison aggregation method when they are planning an activity that is dependent on the weather, yet there are also some participants who prefer to use direct comparison aggregation method all the time.

7.2 Conclusions

Overall we have came up with the following conclusions: When dealing with uncertain data, using aggregation of multiple sources of uncertain data increases the confidence of the users and is more preferred compared to single source independent of how the data is represented. Uncertainty is difficult to communicate and in the case of weather forecasts uncertainty is mostly not presented to the end user, thus the users try to workaround this issue by evaluating multiple sources themselves. Our application makes it easy to workaround the uncertainty by using aggregations. In the context of weather forecasts, range aggregation method combined with the pictogram representation method is the most preferred way to display aggregation of multiple sources. When the user needs more detailed information about how the aggregation was calculated, providing an option to see individual sources by using direct comparison aggregation method is needed.

7.3 Future Work

In future work, these aggregation methods could be applied other contexts that deal with uncertain data such as arrival time predictions, contradicting sensor measurements, or health predictions. To further continue with the context of weather forecasts, studying the results of a long term usage might provide interesting findings.

Bibliography

[BAL12] K. Brodlie, R. Allendes Osorio, A. Lopes. "A Review of Uncertainty in Data Visualization." In: *Expanding the Frontiers of Visual Analytics and Visualization*. Ed. by J. Dill, R. Earnshaw, D. Kasik, J. Vince, C. P. Wong. London: Springer London, 2012, pp. 81–109. ISBN: 978-1-4471-2804-5. DOI: 10.1007/978-1-4471-2804-5_6. URL: http://dx.doi.org/10.1007/978-1-4471-2804-5_6 (cit. on pp. 4, 5).

[Bec87] M. B. Beck. "Water quality modeling: A review of the analysis of uncertainty." In: *Water Resources Research* 23.8 (1987), pp. 1393–1442. ISSN: 1944-7973. DOI: 10.1029/WR023i008p01393. URL: http://dx.doi.org/10.1029/WR023i008p01393 (cit. on p. 3).

[BKM09] A. Bangor, P. Kortum, J. Miller. "Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale." In: *J. Usability Studies* 4.3 (May 2009), pp. 114–123. ISSN: 1931-3357. URL: http://dl.acm.org/citation.cfm?id=2835587.2835589 (cit. on p. 49).

[Bra16] M. Brady. "J M Keynes on the Definition of Uncertainty: Why Uncertainty must come in Degrees and has nothing to do with Ergodicity or Non Ergodicity." In: *Scholedge International Journal of Management & Development ISSN 2394-3378* 3.1 (2016), pp. 1–10. ISSN: 2394-3378. URL: http://thescholedge.org/index.php/sijmd/article/view/270 (cit. on p. 3).

[Bro96] J. Brooke. "SUS-A quick and dirty usability scale." In: *Usability evaluation in industry* 189.194 (1996), pp. 4–7 (cit. on p. 47).

[Cen05] N. S. F. Center. "Weather Forecasting Through the Ages." In: *Water Encyclopedia*. John Wiley & Sons, Inc., 2005. ISBN: 9780471478447. DOI: 10.1002/047147844X.me2114. URL: http://dx.doi.org/10.1002/047147844X.me2114 (cit. on p. 1).

[GLSS00] E. L. Glaeser, D. I. Laibson, J. A. Scheinkman, C. L. Soutter. "Measuring trust." In: *Quarterly Journal of Economics* (2000), pp. 811–846 (cit. on p. 5).

[GRM+08] J. Gill, J. Rubiera, C. Martin, I. Cacic, K. Mylne, C. Dehui, G. Jiafeng, T. Xu, M. Yamaguchi, K. Foamouhoue, E. Poolman, J. Guiney, H. Kootval. “Guidelines on communicating forecast uncertainty.” In: *World Meteorological Organization/Technical Document 4122* (2008) (cit. on p. 4).

[HLHG00] U. Hoffrage, S. Lindsey, R. Hertwig, G. Gigerenzer. “Communicating Statistical Information.” In: *Science* 290.5500 (2000), pp. 2261–2262. ISSN: 0036-8075. DOI: 10.1126/science.290.5500.2261. eprint: http://science.sciencemag.org/content/290/5500/2261. URL: http://science.sciencemag.org/content/290/5500/2261 (cit. on p. 4).

[HP07] J. Handmer, B. Proudley. “Communicating uncertainty via probabilities: The case of weather forecasts.” In: *Environmental Hazards* 7.2 (2007), pp. 79–87 (cit. on p. 4).

[HS13] P. O. Hanrahan, C. Sweeney. “Odds on weather: probabilities and the public.” In: *Weather* 68.9 (2013), pp. 247–250. ISSN: 1477-8696. DOI: 10.1002/wea.2137. URL: http://dx.doi.org/10.1002/wea.2137 (cit. on p. 4).

[Hub10] D. Hubbard. *How to Measure Anything: Finding the Value of Intangibles in Business*. Wiley, 2010. ISBN: 9780470625699. URL: https://books.google.de/books?id=CBAh4eM-g3AC (cit. on p. 3).

[IM87] H. Ibrekk, M. G. Morgan. “Graphical Communication of Uncertain Quantities to Nontechnical People.” In: *Risk Analysis* 7.4 (1987), pp. 519–529. ISSN: 1539-6924. DOI: 10.1111/j.1539-6924.1987.tb00488.x. URL: http://dx.doi.org/10.1111/j.1539-6924.1987.tb00488.x (cit. on p. 5).

[JPJ+07] S. Joslyn, K. Pak, D. Jones, J. Pyles, E. Hunt. “The Effect of Probabilistic Information on Threshold Forecasts.” In: *Weather and Forecasting* 22.4 (2007), pp. 804–812. DOI: 10.1175/WAF1020.1. eprint: http://dx.doi.org/10.1175/WAF1020.1. URL: http://dx.doi.org/10.1175/WAF1020.1 (cit. on p. 4).

[LH99] I. Lipkus, J. Hollands. “The visual communication of risk.” In: *Journal of the National Cancer Institute. Monographs* 25 (1999), 149â€”163. ISSN: 1052-6773. URL: http://europepmc.org/abstract/MED/10854471 (cit. on p. 4).

[Mac92] A. MacEachren. “Visualizing Uncertain Information.” In: *Cartographic Perspectives* 13 (1992). URL: http://www.cartographicperspectives.org/index.php/journal/article/view/cp13-maceachren (cit. on p. 4).

[MDL08] R. E. Morss, J. L. Demuth, J. K. Lazo. "Communicating Uncertainty in Weather Forecasts: A Survey of the U.S. Public." In: *Weather and Forecasting* 23.5 (2008), pp. 974–991. DOI: 10.1175/2008WAF2007088.1. eprint: http://dx.doi.org/10.1175/2008WAF2007088.1. URL: http://dx.doi.org/10.1175/2008WAF2007088.1 (cit. on pp. 3, 4).

[MLD10] R. E. Morss, J. K. Lazo, J. L. Demuth. "Examining the use of weather forecasts in decision scenarios: results from a US survey with implications for uncertainty communication." In: *Meteorological Applications* 17.2 (2010), pp. 149–162 (cit. on p. 5).

[MRH+05] A. M. MacEachren, A. Robinson, S. Hopper, S. Gardner, R. Murray, M. Gahegan, E. Hetzler. "Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know." In: *Cartography and Geographic Information Science* 32.3 (2005), pp. 139–160. DOI: 10.1559/1523040054738936. eprint: http://dx.doi.org/10.1559/1523040054738936. URL: http://dx.doi.org/10.1559/1523040054738936 (cit. on p. 3).

[Nic99] N. Nicholls. "Cognitive Illusions, Heuristics, and Climate Prediction." In: *Bulletin of the American Meteorological Society* 80.7 (1999), pp. 1385–1397. DOI: 10.1175/1520-0477(1999)080<1385:CIHACP>2.0.CO;2. eprint: http://dx.doi.org/10.1175/1520-0477(1999)080<1385:CIHACP>2.0.CO;2. URL: http://dx.doi.org/10.1175/1520-0477(1999)080%3C1385:CIHACP%3E2.0.CO;2 (cit. on p. 4).

[NJ09] L. Nadav-Greenberg, S. L. Joslyn. "Uncertainty Forecasts Improve Decision Making Among Nonexperts." In: *Journal of Cognitive Engineering and Decision Making* 3.3 (2009), pp. 209–227. DOI: 10.1518/155534309X474460. eprint: http://edm.sagepub.com/content/3/3/209.full.pdf+html. URL: http://edm.sagepub.com/content/3/3/209.abstract (cit. on p. 4).

[OM] C. Olston, J. D. Mackinlay. "Visualizing Data with Bounded Uncertainty." In: (), pp. 1–8 (cit. on pp. 4, 5, 14, 16).

[Pal00] T. N. Palmer. "Predicting uncertainty in forecasts of weather and climate." In: *Reports on Progress in Physics* 63.2 (2000), p. 71. URL: http://stacks.iop.org/0034-4885/63/i=2/a=201 (cit. on p. 3).

[PST+13] F. Pappenberger, E. Stephens, J. Thielen, P. Salamon, D. Demeritt, S. J. van Andel, F. Wetterhall, L. Alfieri. "Visualizing probabilistic flood forecast information: expert preferences and perceptions of best practice in uncertainty communication." In: *Hydrological Processes* 27.1 (2013), pp. 132–146. ISSN: 1099-1085. DOI: 10.1002/hyp.9253. URL: http://dx.doi.org/10.1002/hyp.9253 (cit. on p. 5).

[Rai00] M. A. Raimondo. “The measurement of trust in marketing studies: a review of models and methodologies.” In: *16th IMP-conference, Bath, UK*. Citeseer. 2000 (cit. on p. 5).

[Rat07] B. C. Rathbun. “Uncertain about Uncertainty: Understanding the Multiple Meanings of a Crucial Concept in International Relations Theory.” In: *International Studies Quarterly* 51.3 (2007), pp. 533–557. ISSN: 1468-2478. DOI: 10.1111/j.1468-2478.2007.00463.x. URL: http://dx.doi.org/10.1111/j.1468-2478.2007.00463.x (cit. on p. 3).

[RBKS06] M. S. Roulston, G. E. Bolton, A. N. Kleit, A. L. Sears-Collins. “A Laboratory Study of the Benefits of Including Uncertainty Information in Weather Forecasts.” In: *Weather and Forecasting* 21.1 (2006), pp. 116–122. DOI: 10.1175/WAF887.1. eprint: http://dx.doi.org/10.1175/WAF887.1. URL: http://dx.doi.org/10.1175/WAF887.1 (cit. on pp. 4, 5).

[RK09] M. S. Roulston, T. R. Kaplan. “A laboratory-based study of understanding of uncertainty in 5-day site-specific temperature forecasts.” In: *Meteorological Applications* 16.2 (2009), pp. 237–244. ISSN: 1469-8080. DOI: 10.1002/met.113. URL: http://dx.doi.org/10.1002/met.113 (cit. on p. 4).

[SP11] J. Slingo, T. Palmer. “Uncertainty in weather and climate prediction.” In: *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 369.1956 (2011), pp. 4751–4767. ISSN: 1364-503X. DOI: 10.1098/rsta.2011.0161. eprint: http://rsta.royalsocietypublishing.org/content/369/1956/4751.full.pdf. URL: http://rsta.royalsocietypublishing.org/content/369/1956/4751 (cit. on p. 3).

[Sta73] L. Standing. “Learning 10000 pictures.” In: *Quarterly Journal of Experimental Psychology* 25.2 (1973). PMID: 4515818, pp. 207–222. DOI: 10.1080/14640747308400340. eprint: http://dx.doi.org/10.1080/14640747308400340. URL: http://dx.doi.org/10.1080/14640747308400340 (cit. on p. 12).

[SZB+09] J. Sanyal, S. Zhang, G. Bhattacharya, P. Amburn, R. Moorhead. “A User Study to Compare Four Uncertainty Visualization Methods for 1D and 2D Datasets.” In: *IEEE Transactions on Visualization and Computer Graphics* 15.6 (Nov. 2009), pp. 1209–1218. ISSN: 1077-2626. DOI: 10.1109/TVCG.2009.114. URL: http://dx.doi.org/10.1109/TVCG.2009.114 (cit. on pp. 4, 5).

[WFGH11] J. O. Wobbrock, L. Findlater, D. Gergle, J. J. Higgins. “The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures.” In: *Proceedings of the SIGCHI Conference on Human Factors in*

Computing Systems. CHI '11. Vancouver, BC, Canada: ACM, 2011, pp. 143–146. ISBN: 978-1-4503-0228-9. DOI: 10.1145/1978942.1978963. URL: http://doi.acm.org/10.1145/1978942.1978963 (cit. on p. 27).

[WHR+03] W. Walker, P. Harremoës, J. Rotmans, J. van der Sluijs, M. van Asselt, P. Janssen, M. K. von Krauss. "Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support." In: *Integrated Assessment* 4.1 (2003), pp. 5–17. DOI: 10.1076/iaij.4.1.5.16466. eprint: http://dx.doi.org/10.1076/iaij.4.1.5.16466. URL: http://dx.doi.org/10.1076/iaij.4.1.5.16466 (cit. on p. 3).

[YY94] T. Yamagishi, M. Yamagishi. "Trust and commitment in the United States and Japan." In: *Motivation and emotion* 18.2 (1994), pp. 129–166 (cit. on p. 5).

All links were last followed on September 29, 2016.

Declaration

I hereby declare that the work presented in this thesis is entirely my own and that I did not use any other sources and references than the listed ones. I have marked all direct or indirect statements from other sources contained therein as quotations. Neither this work nor significant parts of it were part of another examination procedure. I have not published this work in whole or in part before. The electronic copy is consistent with all submitted copies.

---

place, date, signature