The Effectiveness of Visualization Techniques on Mobile Devices for Time Dependent Data

CS448B Final Project Paper

Pakapark Nik Bhumiwat
Stanford University
pakapark@cs.stanford.edu

Albert Feng
Stanford University
alfeng6@cs.stanford.edu

ABSTRACT

In general, contemporary data visualization research for time dependent data focuses on the accuracy (how accurate an observer understands the data) and efficiency (how fast an observer understands the data) of two techniques: animation and small multiple. However, none of this research takes in account the screen space requirements in visualizing these two techniques. In this paper, we will address and experiment similar techniques (slider graph and small multiple) while considering screen space. Our experiment yields a surprising result that, with 95% confidence, slider graph is significantly more accurate than small multiple to convey the numerical detail of the data in mobile device.

ACM Reference format:

Pakapark Nik Bhumiwat and Albert Feng. 2017. The Effectiveness of Visualization Techniques on Mobile Devices for Time Dependent Data. In *Proceedings of ACM Conference, Washington, DC, USA, July 2017 (Conference'17),* 6 pages.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Most visualization research conducted today do so without consideration for the amount of space in which the visualization takes up. Thus, the results that are extrapolated can only be trusted on devices that do not have screen space constraints. However, as mobile devices become the ubiquitous manner in which people communicate and since visualizations have become a pristine solution for showing off lots of data in a smaller amount of space (in comparison to thousands of rows of data), it becomes imperative to research the effectiveness of popular visualization techniques on mobile devices. Although phone screens are getting bigger, they are still several times smaller than personal computers. For example, the iPhone X has a height of 5.65 inches and width of 2.79 inches, which equates to roughly 17.36 square inches of screen real estate. In contrast, a 15.4 inch Macbook Pro has a screen with a height of 8.2 inches and width of 13.1 inches, which equates to roughly 107.42 square inchesâĂŤ more than six times the screen size as the iPhone X. Therefore, we wanted to understand the effectiveness of visualization techniques on mobile devices.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

2 RELATED WORK AND REFERENCE

Most of the literature focuses on comparing and measuring visualization techniques on devices where screen space is not the issue. "Animation, Small Multiples," explores the effectiveness of small multiple and animation by measuring the accuracy and efficiency at which its participants were able to gauge the number of extra nodes display in one graph when compared to another [3]. This paper concluded that small multiple was a significantly faster means of interpreting data while animation was more accurate for interpreting data.

As for literature on the effectiveness of mobile visualizations, the only literature available for mobile visualizations consist of tips for how to make a good visualization, rather than scientifically tested visualization techniques [2]. Specifically, the most popular articles are those written by major players in the visualization space (i.e. Tableau) and are written for a community of visualization creators rather than visualization researchers. In the article written by Tableau, it speaks to tips such as locking your pan and zoom so you will know where you are in your data and to use range sizing so that your visualization will look differently on different sized devices. Unfortunately, these articles do not mention the specific techniques that should be employed for effective visualizations.

Finally, we wanted to explore visualization techniques on mobile devices for surfacing time dependent data due to the extra layer of complexity associated with time affecting the data. This extra dimension lends itself to data that requires extra level of interpretation from the user, making the differences in effectiveness of the visualization techniques much clearer. As for related work in this space for time dependent data, most research has been on the complexity of visualizing lots of data associated with time dependent data. "Visualizing Time-Oriented Data-A Systematic View" goes into depth on the types of time-oriented visualizations: linear, cyclic and branch and presents a high level overview of the various types of graphs that can be used [1]. In this paper, Aigner focuses on the matter that because there is so much more data that needs to be surfaced for time-oriented data, the creator of the visualization must think on first principlesâĂŤspecifically on the characteristics of data that must be shown. From these papers, we came to the conclusion that we would focus on linear time dependent data because of the many other factors we would need to consider for cyclic or branch visualizations. Furthermore, we would need to orient our experiment with data that we generate ourselves so that we can control the amount of data that needs to be surfaced and the parts of the data that would be significant to the user.

3 METHODS

We wanted to narrow our exploration and assessment of visualization techniques on mobile devices by surfacing time dependent data with two visualization techniques (slider graph and small multiple) using two visualization types (double bar graph and scatter plot) to ask two types of questions (detail and trend) in order to measure two main factors (accuracy and efficiency).

A small multiple is a series of similar graphs/charts where the data can be more easily compared and a slider graph is a single graph where the data shown changes depending on where the slider is positioned. These two visualization techniques are commonly found for comparing time slices and provide an interesting combination for screen size constrained visualizations since each individual small multiple time slice is much smaller than the one graph shown in the slider graph. In addition, there is also the comparison between one visualization technique that surfaces all the data at once (small multiple) and another visualization technique that surfaces individual time slices while hiding the others (slider graph). As for the types of graphs shown, we decided to use double bar graphs and scatter plots because they are two of the most commonly used graphs and help show off different aspects of the data. Scatter plots help show trends from individual data points while double bar graphs do a better job of comparing between two entities.

For the types of questions we ask, we decided upon multiple choice questions that ask for either trend analysis or detail analysis since these are two of the most common types of analyses users of visualizations would have when approaching a visualization. Multiple choice questions allow for us to easily create more generalizable questions and gauge the correctness of the questions we ask as well as have questions with only limited answer choices to prevent participants from being overwhelmed. A trend question would consist of comparing two entities and examining how its data points change in relation to time. A detail question would consist of identifying specific data points within two entities and computing the absolute difference between those points. Finally, we decided to measure accuracy and efficiency because those are the two main attributes that make for an effective visualization. To measure accuracy, we record whether the participant answered the question correctly and which question the participant answered. To measure efficiency, we record the amount of time it takes for the participant to answer any given question.

4 TESTING PLATFORM AND EXPERIMENT

4.1 Testing Environment

We developed a mobile testing platform with standard web application frameworks using a MEN stack (MongoDB, Express, and Node.js) where the front end was developed with embedded JavaScript, also known as *ejs*. This program was hosted in Heroku and allowed for easy remote access through an encrypted URL: bit.ly/viz-mobile-research. This program was also designed to allow users to access the link only via mobile phones as shown in Figure 1.

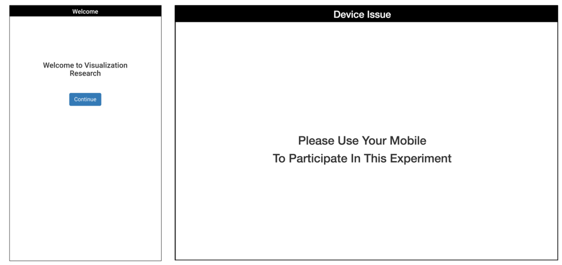

Figure 1: Testing Environment is limited to mobile platform only. Home page for data collecting page on mobile (left) and on desktop (right).

4.2 Blind Test

Before any participants engaged in our experiment, we asked for a consent form to guarantee that the information we collect would only be used for research and would not be revealed in general public. In addition, we assigned each participant with a random 7-digit number to make sure that all the tests are conducted securely. Once the participant agreed to the form, we immediately collected basic information including screen width, screen height, and pixel ratio of the mobile screen. The consent form screen is as shown in Figure 2.

4.3 Data Generator

In order to make sure that we appropriately measured the accuracy and efficiency of each type of visualization, we designed the data generator to output data in the same format for all types of visualizations. For each data set, it contained nine time slices, each of which had ten data points (five of them are labelled as A and the other are labelled as B). Each data point was a real number ranging from 0 to 5. In the data randomization process, we made sure that there was a clear relationship between data in group A and B by randomly selecting one of the following possible categories:

(1) A > B and all data points increase over time
(2) A < B and all data points increase over time.
(3) A ∼ B and all data points increase over time.
(4) A > B and all data points decrease over time
(5) A < B and all data points decrease over time.
(6) A ∼ B and all data points decrease over time.
(7) Data in A increases over time but data in B is decreases over time.
(8) Data in A decreases over time but data in B increases over time.

where

(1) $A > B$ means all data in A is always greater than data in B.
(2) $A < B$ means all data in A is always smaller than data in B.
(3) $A \sim B$ means all data in A is equivalent to all data in B.

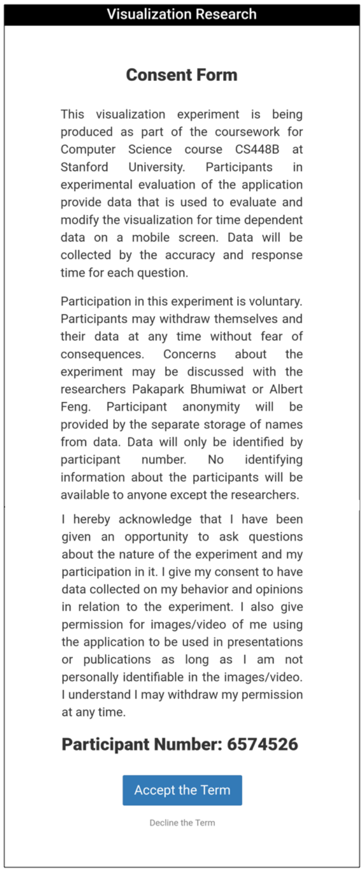

Figure 2: Data collected from participants are labelled with participant number both for prevent data distortion and transparency of the experiment.

To make sure that this data would not be altered until it was delivered to each participant, all the data was generated from the server side only when it received a request from a local machine with a hidden verified ID that was generated after the participant had agreed with our consent form.

4.4 Questions

As explained in part 3, we asked participants two warm up questions and eight actual questions. The two warm up questions were predetermined to be a slider graph/double bar graph/trend question and a small multiple/scatter plot/detail question so that participants could familiarize themselves with the two visualization techniques, two visualization types and two question types. An example of the warm up questions is shown in Figure 3.

The eight actual questions were the eight possible combinations of visualization techniques, types of graph, and questions in random order. As for a specific implementation for trend questions, we asked the participant to select the relationship between four multiple

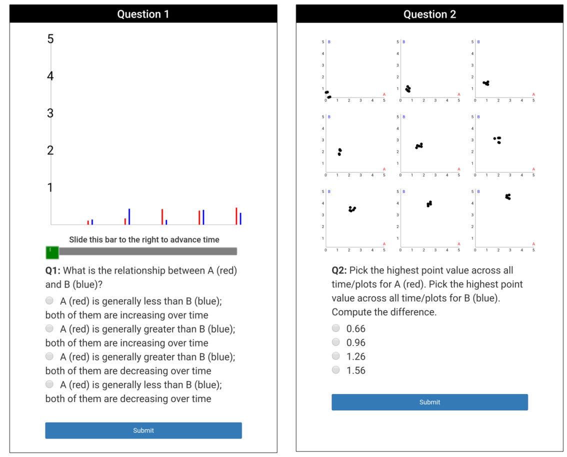

Figure 3: Example of two warm up questions

choices that were randomly generated corresponding to the type of data described in the previous part. (Note that we also made sure that all questions did not have ambiguous answers and seamlessly formatted them in an order that made sense to all participants.) An example of a trend question can be seen in Figure 4-5.

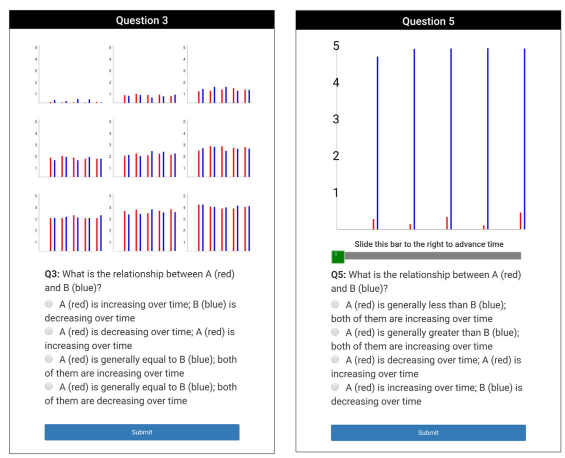

Figure 4: Example of two trend questions with bar plot: small multiple (left) and slider graph (right)

For detail questions, we asked participants to determine the difference between the largest value of data in group A at any time slices and the largest value of data in group B at any time slices. The multiple choices are formatted in a way that each multiple choice is a nonnegative real number that has a 0.3 difference from each other. An example of a detail question is shown in Figure 7-10.

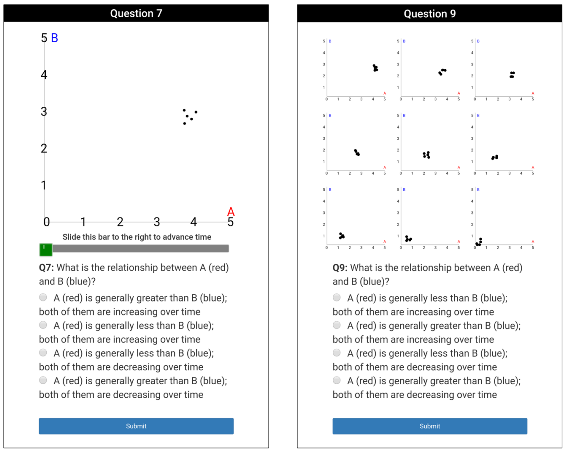

Figure 5: Example of two trend questions with scatter plot: slider graph (left) and small multiple (right)

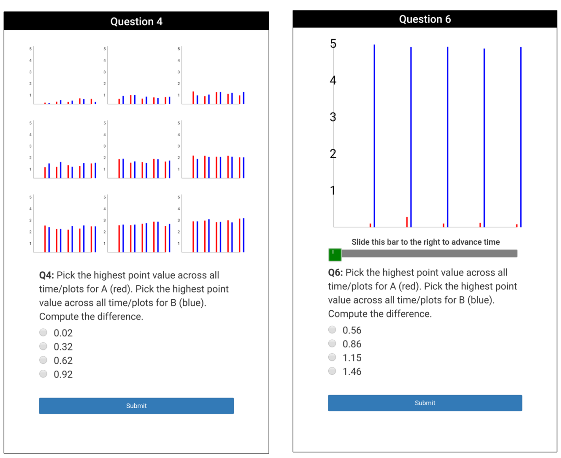

Figure 6: Example of two detail questions with bar plot: small multiple (left) and slider graph (right)

4.5 Data Collection

To compare the accuracy and efficiency of small multiple and slider graph as well as find out any other possible correlations, we collected a handful of information as shown in Table 1.

1. **Correctness**: A binary result for each question (correct or wrong)
2. **First Decision Time**: The amount of time that each participant spends before deciding to choose an option
3. **Reluctance**: The number of time that participant changes the answer
4. **Last Decision Time**: The amount of time that participant spends for deciding the last option before submission

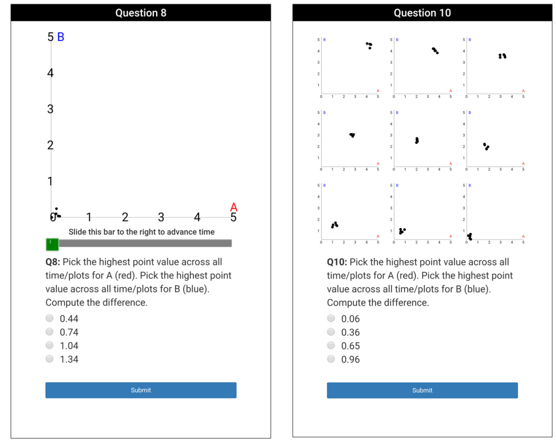

Figure 7: Example of two detail questions with scatter plot: slider graph (left) and small multiple (right)

5. **Question Order**: The order of question that each participant has in the experiment
6. **Screen Width**: The screen width of the participant's mobile device
7. **Screen Height**: The screen height of the participant's mobile device
8. **Pixel Ratio**: The screen pixel ration of the participant's mobile device
9. **Gender**: Participant's gender

5 RESULT

In this part, we provide aggregate data from 16 participants for each variable described earlier. Note that the data provided in this part is intended to only highlight the trend of our data. We will provide in depth analysis in the discussion section.

Variable	Data Collected
Gender	12 male participants, 1 female participant, and 3 participants would like to not specify their gender
Reluctance	Almost all participants does not change their answers for almost all questions.
Question Order	Two participants per each question order. (8 possible question orders in total)
Screen Width	Range from 320 to 414 pixels
Screen Height	Range from 568 to 846 pixels
Pixel Ratio	Range from 2 to 4
Device	All participants complete the experiment with mobile phone

Figure 8: Aggregated basic information about participants

5.1 Correctness

We define correctness as a binary variable such that for each question it will have a value of 1 if and only if the participant select the right answer. Otherwise, it will have a value of 0. The aggregate data from all participants can be found in Figure 9.

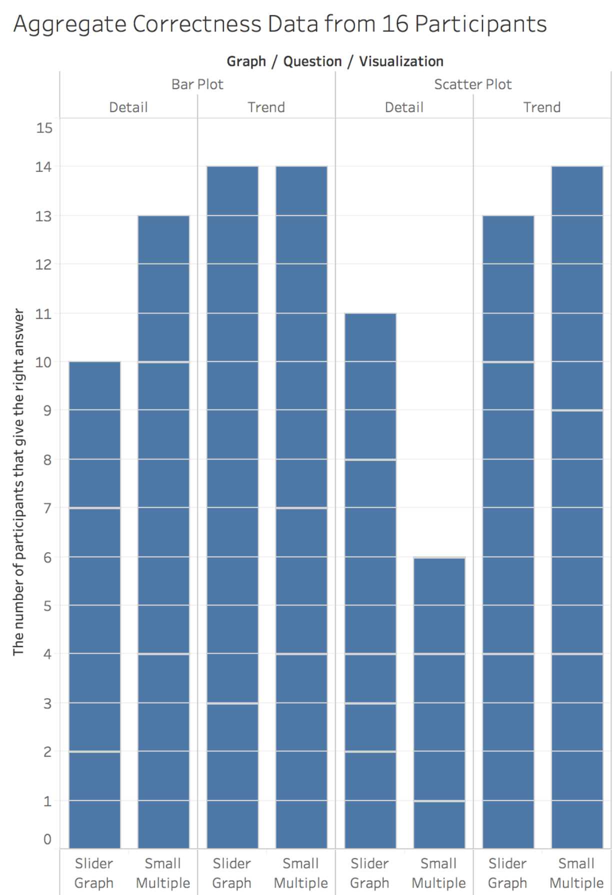

Figure 9: Aggregated correctness data from 16 participants for each type of visualization, each type of graph, and each type of questions

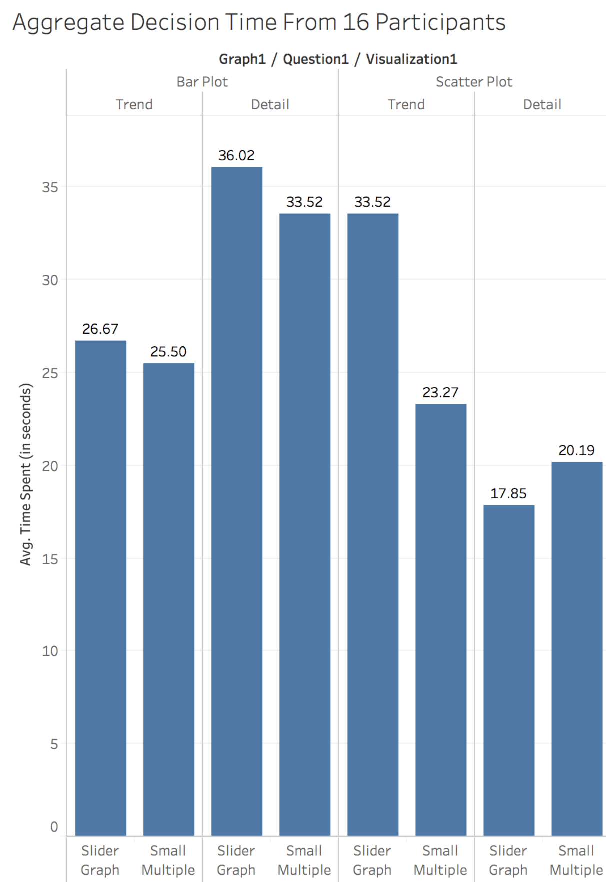

Figure 10: Average first decision time from 10 participants for each type of visualization, each type of graph, and each type of questions

5.2 Decision Time

Decision time is the amount of time it takes for the participant to first decide upon a selected answer. Because almost all participants do not change their options during the experiment, the values for first decision time and last decision time are almost equivalent for all questions. We use this decision time for further analysis and the aggregate data can be seen in Figure 10.

6 DISCUSSION

To compare the accuracy (correctness) and efficiency (decision time) between slider graph and small multiple, we applied one-tailed pair difference t-test under four scenarios:

Case I: Slider Graph vs Small Multiple for Bar Graph
Case II: Slider Graph vs Small Multiple for Scatter Plot
Case III: Slider Graph vs Small Multiple for Trend Questions
Case IV: Slider Graph vs Small Multiple for Detail Questions

The p values for both accuracy and efficiency of each visualization comparison under four scenarios are shown in Figure 11.

6.1 Accuracy

From Figure 11, we can see that these two visualization techniques are not significantly different for accuracy in case I-III. However, in case IV, we find that with 95% confidence, slider graph significantly yields more accuracy than small multiple for numerical details in mobile devices.

6.2 Efficiency

From Figure 11, we notice that there is no significant difference between these two visualization in terms of efficiency (i.e. decision time). After looking into more details in each participant's data, we find out that this might be an error within the subject. Even though we told each participant to complete the experiment as fast as possible while perform the best correctness, the experiment is conducted without in-person monitor making the time extremely variable between each person depending on the environment they completed the experiment. We believe that the in-person monitoring is required in order to have more accurate evaluation for the efficiency of the visualization.

p-value		
Case I: Slider Graph vs Small Multiple for Bar Graph		
Correctness	Slider Graph > Small Multiple	0.2860
Decision time	Slider Graph > Small Multiple	0.1200
Case II: Slider Graph vs Small Multiple for Scatter Plot		
Correctness	Slider Graph > Small Multiple	0.1055
Decision Time	Slider Graph > Small Multiple	0.2605
Case III: Slider Graph vs Small Multiple for Trend Questions		
Correctness	Small Multiple > Slider Graph	0.3309
Decision Time	Slider Graph > Small Multiple	0.1973
Case IV: Slider Graph vs Small Multiple for Detail Questions		
Correctness	Slider Graph > Small Multiple	0.0449
Decision Time	Slider Graph > Small Multiple	0.1634

Figure 11: p-values from one-tailed pair difference t-test comparing two types of visualizations for both correctness and decision time under four scenarios. Note that A > B for correctness means participants responded more correctly in visualization A than in visualization B. Also, A > B for decision time means participants took a longer time to understand visualization A than visualization B.

6.3 Other Dependencies

Even though the variation in mobile screen sizes is not a major focus in this paper, the regression test suggests that there is no direct relationship between the area of different sized mobile screens and either accuracy and efficiency for each visualization where the area of mobile screen is defined as

$$\text{area} = \text{screen width in pixel} \times \text{screen height in pixel} \times \text{pixel ratio}^2$$

Thus, this paper does not conclude any absolute result in this aspect since we believe our data is too sparse for real numerical analysis for screen size.

In aspect of gender, our data collection is slightly skewed as the majority of the participants are male. We hope that we could recruit for more female participants to reach the conclusive summary whether there exists a correlation between gender and the perception of each type of visualization technique on mobile screen.

7 FUTURE WORK

This research lays the groundwork for further experimentation on the major differences in visualization design and techniques for mobile devices when compared to traditional devices with much larger screen sizes.

Based on our finding that slider graph allowed for more accurate interpretation for detail questions than small multiple, we would love to explore the exact break point in visualization size where data becomes undecipherable. Although we collected data on the screen sizes of our participants, there was not enough range of sizes for us to extrapolate anything meaningful in this particular experiment. This would require us to show a specific visualization type and technique with various sizes and have participants answer questions associated with that visualization.

We would also like to explore the differences in accuracy and efficiency between mobile devices and computers rather than simply exploring the effectiveness of techniques within the collection of mobile devices. Although there is research already conducted on the effectiveness of visualization techniques on computers, we would need to conduct an experiment where the conditions for mobile and computers would be the same so that the data inferred can be useful.

Finally, because the manner in which we interact with mobile devices (ie. touch and gestures) is inherently different from that of the computer (ie. keyboard and mouse), there is a significant amount of research to be conducted on the effectiveness of the mechanisms for using mobile devices. Especially with larger amounts of data, mobile visualizations must be able to pan and zoom into different granularities for the user to gather the information they require. Our experiment only consisted of 90 data points for each visualization but many real world data sets consist of at least 9,000 data points. Thus, it becomes imperative to understand how to sift through so much data in such a small device.

REFERENCES

[1] Wolfgang Aigner. 207. Visualizing Time-Oriented Data âĂŞ A Systematic View. (207). http://www.sciencedirect.com/science/article/pii/S0097849307000611
[2] Stefano Burigat and Luca Chittaro. 2011. On the effectiveness of Overview+Detail visualization on mobile devices. (2011). http://delivery.acm.org/10.1145/2440000/2434670/779_2011_Article_500.pdf?ip=128.12.244.132&id=2434670&acc=ACTIVE%20SERVICE&key=AA86BE8B6928DDC7%2E0AF80552DEC4BA76%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&CFID=838163289&CFTOKEN=40203206&__acm__=1512711067_a52ea04746ccd59380056ab1ffe52098
[3] Helen Purchase Daniel Archambault and Bruno Pinaud. 2010. Animation, Small Multiples, and the Effect of Mental Map Preservation in Dynamic Graphs. (2010). http://ieeexplore.ieee.org/document/5473226/