

What does Data Visualization Reveal about Children's Activities in RoboTutor?

-- Based on Descriptive Data Analysis

Lei Xu

University of California, Los Angeles, Los Angeles, 90095, CA, USA

Abstract. We are in the era of promising mobile education. However, data analysis based on children's activities is still limited. This article used rural Tanzanian children's activity log data in site 131 from November 2018 to January 2019 provided by the RoboTutor team and applied the method of Descriptive Data Analysis to roughly analyze children's activity through data visualization. First, children may be more interested in story and literacy than math activities; second, children generally spend more time on tasks they cannot fully complete; finally, all children spend more than 60% of their time completing activities, usually in the morning or afternoon. Using this result can provide a more intuitive understanding of children's activity patterns, which can be helpful for researchers to measure children's activity engagement.

Keywords: Children's Activities; Robotutor; Descriptive Data Analysis; Data Visualization.

1. Introduction

Children have demonstrated positive learning outcomes as a result of attempts at mobile education technology in certain organizations [2]. RoboTutor, as a tablet-based learning system, can be a key to facilitating children's development of reading and arithmetic skills [3]. Findings revealed from children's activities on RoboTutor may help researchers better identify children's activity patterns and provide better technical support. What characteristics can we find out about children's activities through data visualization? In this work, we use the method of Descriptive Data Analysis to analyze 2139 rows of the Tanzanian children's activity log data from RoboTutor in site 131 from November 2018 to January 2019. We carry out data visualizations on children's activity duration, children's activity time, the percentage of time in activities, children completing activities and story's low 100% completion. Children's activity duration reflects children's patterns of solving problems. Children's activity time and the percentage of time in activities reflect children's activity engagement on RoboTutor. Children completing activities reflect the activity areas kids are interested in. The story's low 100% completion reflects the phenomenon that kids have difficulty selecting story activities. Researchers can use visualized data to identify the characteristics of children's activity patterns and make adjustments to help children learn better.

For the rest of the paper, firstly, the preparatory work before analysis is described in Section2. Then, the analysis process is described in Section3. Section4 acknowledges limitations and future work. Finally, Section5 describes the results of the analysis and concludes.

2. Previous Preparation

The data is based on RoboTutor's activity log data, which includes the children's activities table in site 131 from November 2018 to January 2019. After initial analysis, we encountered some rows with extremely long hiatuses, which could be a miscalculation or kids leaving in the middle. We also discovered the task name of the last attempt of an activity in some rows is null and some rows under content areas are DNFT, which represents incomplete or invalid activity names. We further observed from the table that the ActivityName and TutorID are different for some rows. We consider the data with the back button and crashed status as not completed, but there was no such row to display whether an activity was completed.

To figure out the percentage of rows with null values, we visualize the distribution of the task name of the last attempt of an activity in Fig.1. The horizontal coordinate represents the task name of the last attempt of activities and the vertical coordinate represents their percentage. This chart shows that almost 50% of the task name are back buttons and 1% null values. Back button means that the child pressed back button and left the activity midway.

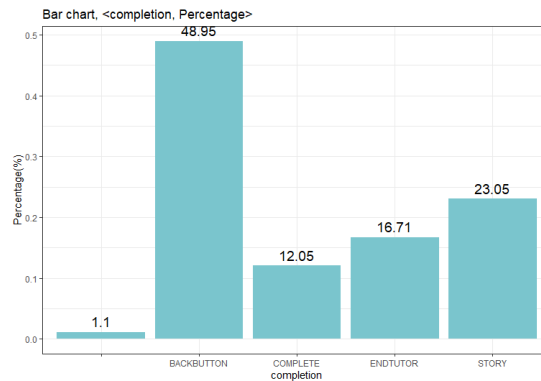


Figure 1. Percentage of Task Name of the Last Attempt

To remove rows with extremely long hiatuses, we winsorized the data by filtering out hiatus outliers with three standard deviation values above. In the chart, we found 1.1% of rows with null values, so we removed these rows. We also excluded the Area_AN column with Did Not Find Tutor (DNFT), because DNFT occurs for select_activity and it is not an actual activity. After that, we filtered out the rows where ActivityName and TutorID are the same. To indicate whether an activity was completed, we added a Completion_2 column to the CSV file. We only consider not completed cases in sections D and E.

3. Analysis Process

3.1 Analysis of Children’s Activity Duration

To analyze children’s activity time, we removed the rows with zero time between the first and last attempt, i.e., rows with zero active duration, and visualized the distribution of the rest of the data. Moreover, since the activity duration time is not intuitive enough, we take the logarithm of the x-axis to make the values equally spaced horizontally. Finally, we get Fig.2, which visualizes the results with x in logarithm scale 5. The horizontal coordinate represents the scaled activity duration and the vertical coordinate represents the count. We see that in logarithm scale 5, the activity duration usually happens between $21558(5^6.2)$ - $205197(5^7.6)$ ms. After unit conversion, we can tell from the visualization that children often spend about half a minute to three and a half minutes answering questions, and they rarely take a few seconds or more than four minutes to answer a question. This indicates that children tend to solve problems for short periods of time.

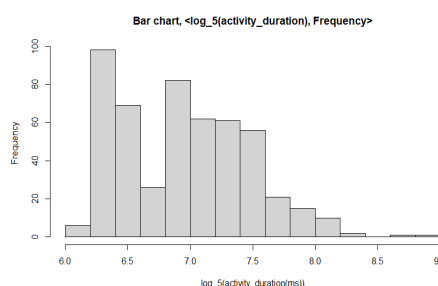


Figure 2. Children’s Activity Duration Distribution

3.2 Analysis of Children’s Activity Time

To find the pattern of children’s activity time, we get the date and time parsed from the name of the PERF log corresponding to the activity and grouped by hours and time of a day, and we consider morning: 6:00-12:00, afternoon: 12:00-18:00, evening: 18:00-24:00, night: 0:00-6:00. We visualized the distribution of activity time grouped by hours and attached percentages accordingly. By analyzing the DateTime_logName grouped by hour shown in Fig.3, we see that most kids do activities during 8-9 am. After grouping by the time of the day, in Table 1, we find the kids usually do activities in the morning and afternoon.

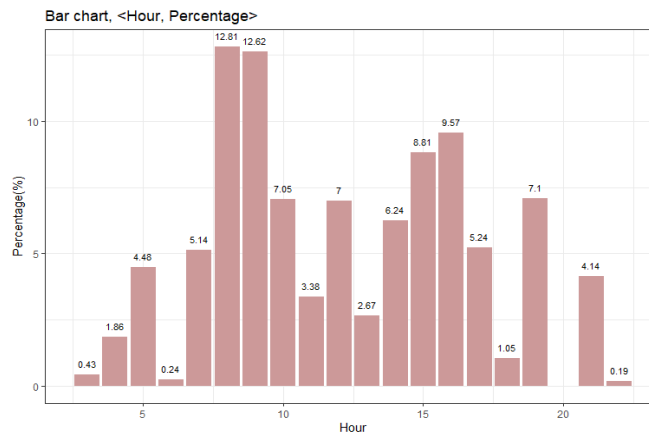


Figure 3. Children’s Activity Time on Hour Basis

Table 1. Children’s Activity Time

Time	count	percentage
morning	176	34.24
afternoon	231	44.94
evening	55	10.70
night	52	10.12

3.3 Analysis of the Percentage of Time in Activities

Based on the child's completion of activities, we visualize how much time the child spends on each activity. Therefore, we made a stacked bar chart of activity duration on different tablets and their completion status.

Fig.4 shows the percentage of children completing activities on different tablets. The horizontal coordinate indicates the percentage of activity duration, the vertical coordinate represents tablet IDs, and the colors are the activity completion status. We can see that kids on three tablets spent more than 60% of their time completing activities, and on one tablet, kids spent almost 90% of their time completing activities.

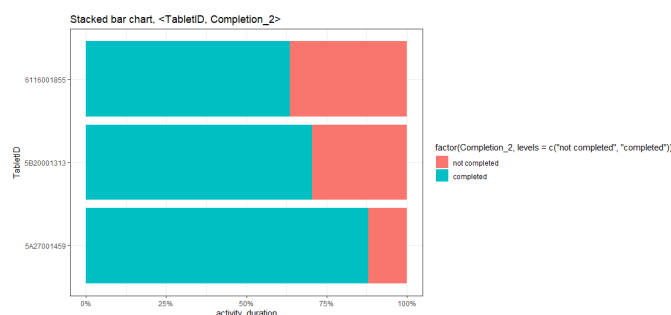


Figure 4. Children’s Activity Duration on TabletID and Completion Basis

3.4 Analysis of Children Completing Activities

To explore the percentage of time that children spend on each content area, we visualize children’s activity completion percentage in different areas with a pie plot in Fig.5. It is seen that among the completed activities, literacy accounts for approximately 50% of the total activities, math accounts for 20% and story accounts for 30%. Since version “cd2” rotated content areas to focus more on literacy, we consider literacy accounts for approximately 25% of the total activities. Children seem less likely to start math problems than literacy and story problems. Mastery-based tests have the potential to increase students' confidence in their mathematical abilities [1].

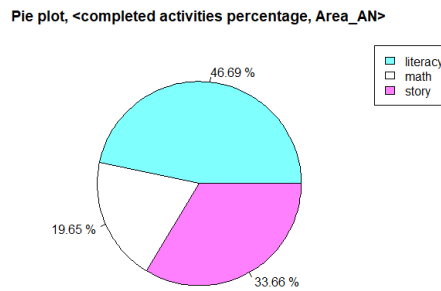


Figure 5. Children’s Activity Completion on Area Basis

We also visualize children’s activity completion percentages in different areas with a scatter plot in Fig.6. The horizontal axis represents the content areas and the vertical axis represents the completion percentage. For this scatter plot, we convey the number of the coincident points by adjusting the point size to the number of identical values. The point with the largest size in each content area is at 100 percent completion. We can tell from the point size that most kids completed 100 percent of their activities.

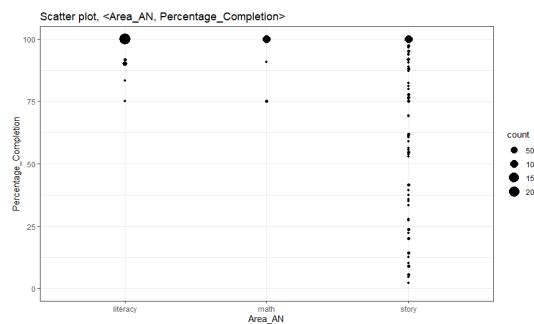


Figure 6. Children’s Activity Completion Percentage of Content Areas

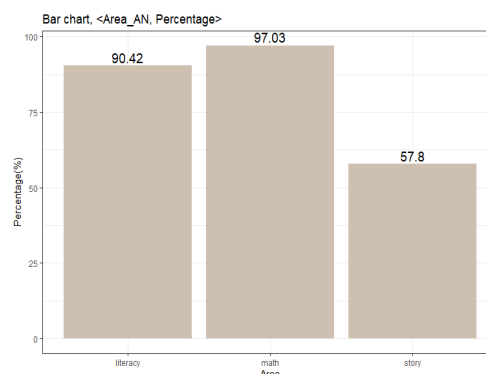


Figure 7. Children’s 100% Completed Activities Count and Percentage

Afterward, we made a visualization of children’s 100% completed activities. It is shown in Fig.7 that 90% of the completed literacy activities are 100% completed and 97% of the completed math activities are 100% completed. Compared to literacy and math, the 100% completion rate of the story is 58%. The 100% completed percentage of stories is far below that of literacy or math.

Table 2 reveals the average and median activity duration of activities with 100% and not 100% completion. It is seen that activities that were not 100% complete had twice the mean and median time than activities that were 100% complete. Kids may spend more time on activities they cannot 100% complete.

Table 2. Children’s Average and Median Activity Duration on 100% Completion Basis

complete		average duration	median duration
100% completed		87469.25	61368
not 100% completed		174576.65	134532

3.5 Analysis of Story’s Low 100% Completion Percentage

In Fig.8, each dot represents one activity split by the number of attempts against activity duration. The horizontal coordinate represents activity duration, the vertical coordinate represents the number of attempts, colors represent content areas, and dot size represents the number of cases. The blue dot in the bottom left records activities with a single attempt and zero activity duration. We learn from the plot that one attempt and zero activity duration in the story area contains the largest number of points. We can tell that kids may be more likely to encounter crashed, back button or DNFT situations in the story area.

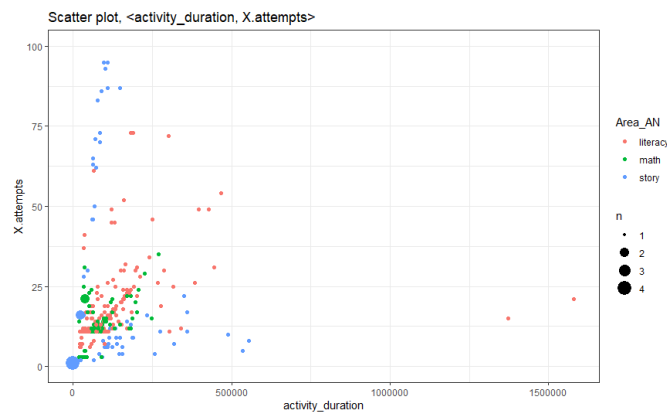


Figure 8. Activity Duration and Number of Attempts

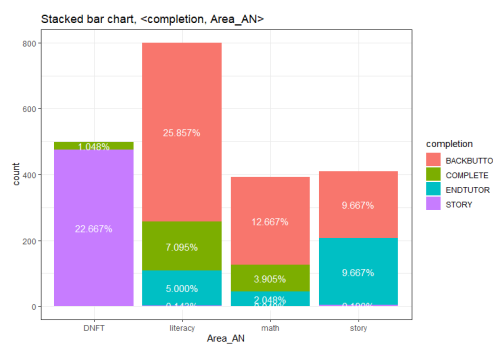


Figure 9. Percentage of Task Name of the Last Attempt

To further verify the story’s low 100% completion phenomenon, we created the stacked bar chart in Fig.9, which displays the task name of the last attempt of an activity in different content areas and their percentages. The horizontal coordinate represents content areas, the vertical coordinate

represents the count, and the colors represent the task name of the last attempt in an activity. DNFT makes up most of the story, therefore under DNFT, there could be some technical issues that impact data logging or prevent kids from selecting activities smoothly. Back button occupies more than half of literacy and math. About half of the story is made up by the back button. We discover that there is a high back button rate under three content areas.

4. Limitations and Future Work

One concern about the findings was that we only analyzed log data from RoboTutor in site 131 from November 2018 to January 2019, which is a small sample size. In addition to this, in section 2, we trimmed half rows with outliers or unwanted values, which further reduces the reliability of the analysis. Therefore, future research should be undertaken with more data to verify and improve the results. Another limitation is that we did not identify why DNFT mostly happens in the story in section F. Considerably, more work will need to be done to figure out what causes this observation.

5. Result and Conclusion

1. Kids generally spend at least 60% of their time on RoboTutor completing activities. They usually started their activities in the morning and afternoon, mainly in the morning during 8-9 am. Therefore, the application maintenance time can be placed in the evening or at night to avoid kids' peak usage.

2. In section E, it can be seen that children prefer to do literacy and story problems rather than math problems. Therefore, children may need mentorship or guidance to intrigue their interest in starting math problems. Data analysis combined with machine learning is promising to demonstrate how many attempts will lead to mastery of a particular problem. However, for math problems that have been started, children can fully complete them with a high percentage.

3. Most children can complete 100% of activity tasks in literacy and math areas but not in the story area. In section H, we see that DNFT is more likely to happen in selecting story activities in RoboTutor data, so the size of valid data in the story is reduced after data cleaning. The low 100% completion rate of the story area may be explained here.

4. In sections B and F, we discover that children often spend half a minute to three and a half minutes answering questions and they may spend twice more time on some activities that they cannot 100% complete. Therefore, people can determine if the difficulty of a question is appropriate for the child's current level and adjust the difficulty of the question based on how much time the child takes to answer it.

References

- [1] Lenarz, J., & Pelatt, K. (2020). A Transition to Mastery-Based Testing with the Hope of Increasing Student Persistence. PRIMUS, 1-10.
- [2] Uchidiuno, J. O. (2020). Facilitating Help-Seeking for K-2 Students in Rural Schools in Tanzania (Doctoral dissertation, Cornell University).
- [3] Zhao, R. (2021). RoboTutor: Predictions of Students' Answer Type. In: Cao, W., Ozcan, A., Xie, H., Guan, B. (eds) Computing and Data Science. CONF-CDS 2021. Communications in Computer and Information Science, vol 1513. Springer, Singapore. https://doi.org/10.1007/978-981-16-8885-0_18.