PlushPal Study

Storytelling with Interactive Stuffed Animals and Machine Learning

Deanna Gelosi, Information Visualization and Presentation (Spring 2021)

Project Goals

PlushPal (https://plushpal.app) is a web app that introduces machine learning (ML) and data science concepts to children through play. Our team[1] hosted workshops and one-on-one play sessions with children ages 8-14 using PlushPal, a micro:bit[2], and a stuffed animal of their choosing. Learners use the web tool[3] to record custom gestures with their stuffed animal and microcontroller to train machine learning models. They then pair their gestures with sounds, which are played back when the model recognizes a gesture trigger.

Through the development of this tool and the facilitation of workshops, the following research questions guided our study:

> **RQ1:** How do children bring their stuffed animals to life using gestures and sound?
> **RQ2:** In what ways did children engage with data science practices when building their ML models with PlushPal?

Though the project wrapped in January 2021 and the findings were submitted to a conference, I wanted to explore other **ways to communicate the research findings in an accessible and playful manner** to researchers familiar with ML and children as well as to practitioners interested in using PlushPal with kids. In this project, I redesigned existing figures and offered new ways of exploring the collected data. The tasks in this visualization are interactive and informative, and aligned with the spirit of learning through play and self-exploration.

Related Work

The following are three informative works that were referenced to the paper we submitted to the ACM Interaction Design and Children conference, each presenting different methods on introducing machine learning to children or makers.

[1] Abigail Zimmermann-Niefield, Shawn Polson, Celeste Moreno, and R Benjamin Shapiro. 2020. Youth making machine learning models for gesture-controlled interactive media. In Proceedings of the Interaction Design and Children Conference.

---

[1] Researchers on this project are from University of Tokyo, Simon Fraser University, and UT Austin.
[2] micro:bit, https://microbit.org.
[3] PlushPal, https://plushpal.app.

Researchers in this paper combined a beginner-level ML modeling toolkit with a beginning programming tool and then studied how young people created and remixed projects to incorporate custom ML-based gestural inputs. Their youth-led projects were more open ended than PlushPal, which specifically used stuffed animals, but the work the children engaged in around custom ML-based gestural inputs is quite similar.

[2] David A Mellis, Ben Zhang, Audrey Leung, and Björn Hartmann. 2017. Machine learning for makers: Interactive sensor data classification based on augmented code examples. In Proceedings of the 2017 Conference on Designing Interactive Systems. 1213–1225.

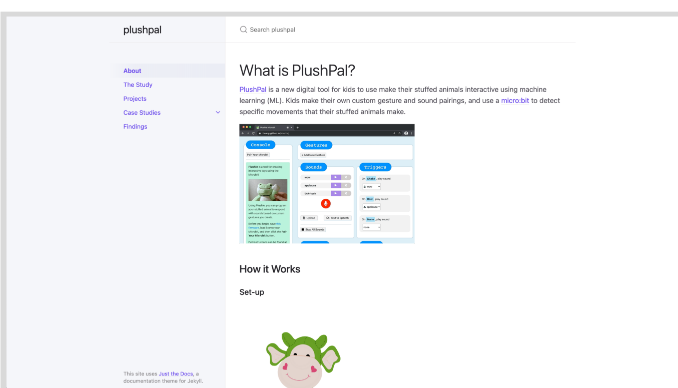
Researchers explore a tool designed for makers to explore machine learning ideas through the support of ML experts. This model supported novices in using ML-powered tools through sensor data collection. Though it's a different audience than that for PlushPal, the hardware and topic similarities made it a compelling example.

[3] Stefania Druga. 2018. Growing up with AI: Cognimates: from coding to teaching machines. M.S. Dissertation. Massachusetts Institute of Technology.

In this Masters thesis, the researcher conducts a study with over one hundred children ages 7 to 14 years old to see if her tool Cognimates helps them develop a better understanding of AI concepts and changes their perspective on AI. The combination of web app tool creation, topic of AI, and the primary audience of children makes this research closely aligned with the PlushPal study.

Visualization

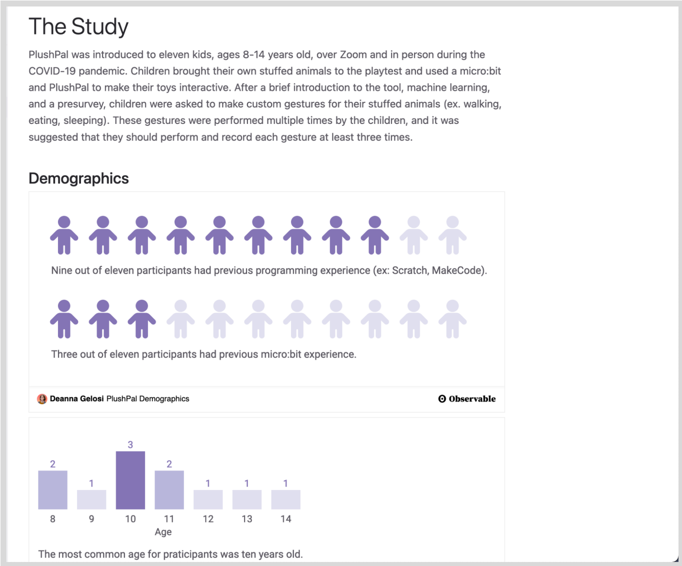
The web page is divided into five sections and contains a combination of images, charts, and interactive visualizations.



About

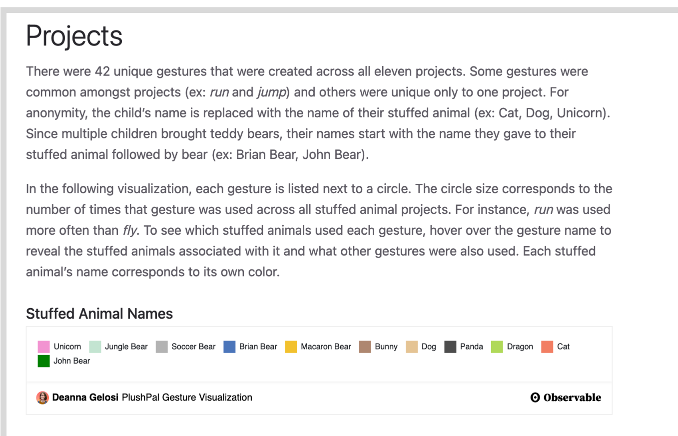
The index page for the PlushPal Study. It includes a description of the tool as well as a side navigation bar to jump to a specific section.

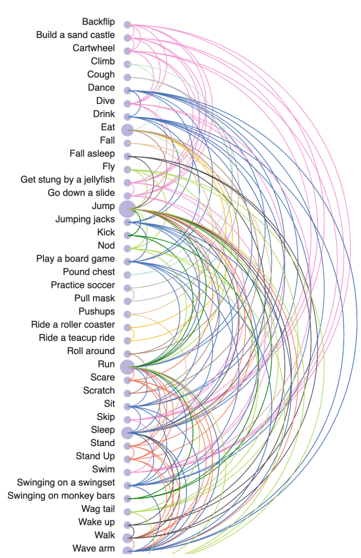
This site uses a Jeckyll theme called 'Just the Docs.'



The Study

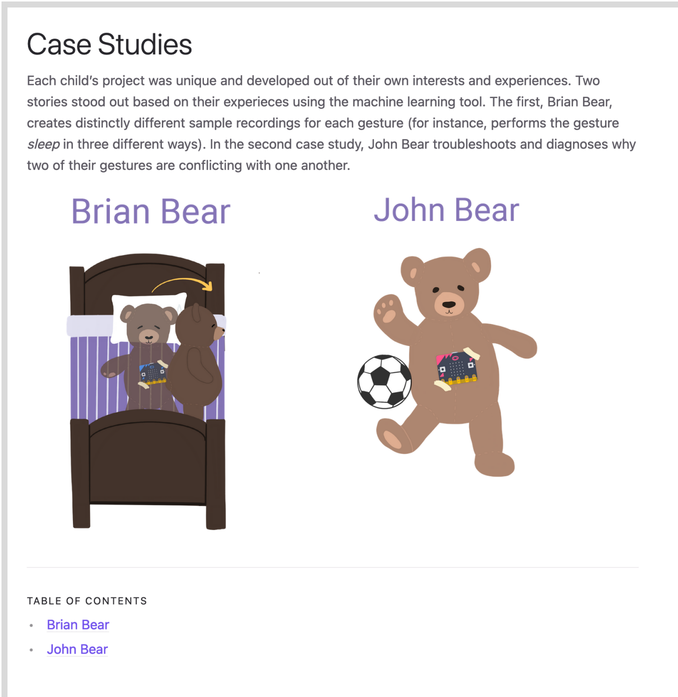
A description of who participated in the study, along with isotypes and a bar chart. These charts were made in D3 using an Observable notebook, and the child icon was created by Musmellow and found on The Noun Project.





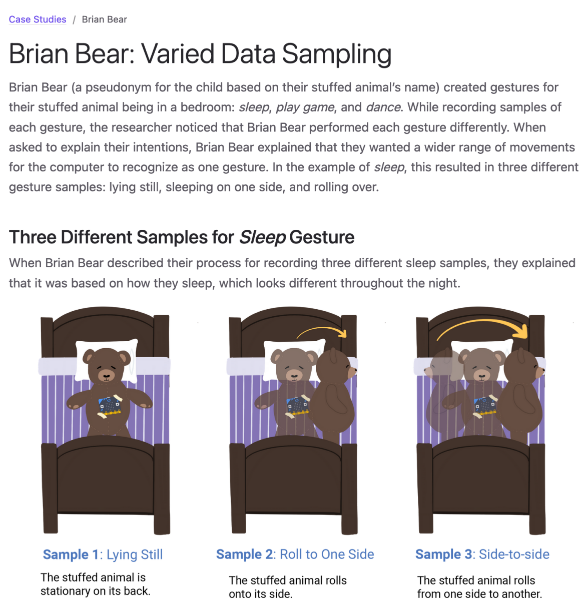
Projects

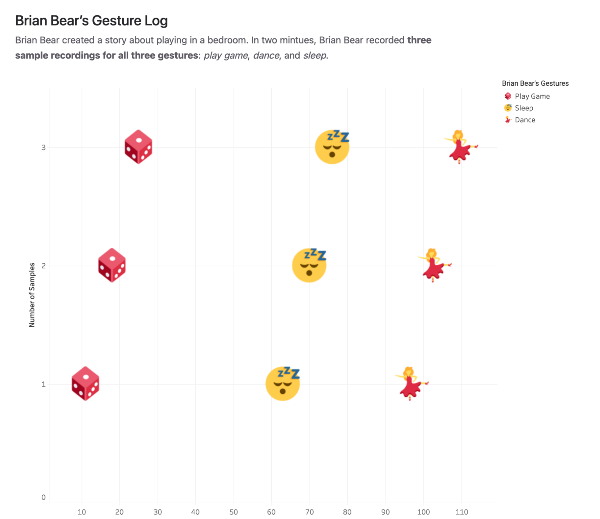
An interactive D3 is a tool to explore all 42 gestures created by children in the study and how they overlap between projects. The interactive visualization was made in an Observable notebook.



Case Studies

The case study landing page introduces both case studies and the images are clickable to jump to their respective child pages.





Brian Bear's case study features an illustrative representation of the three different sleep samples recorded and a log demonstrating their process. The log was made in Tableau and the illustrations were made in Procreate and Illustrator.

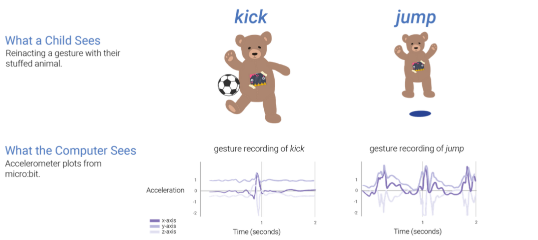
Case Studies / John Bear

John Bear: Indistinguishable Gestures

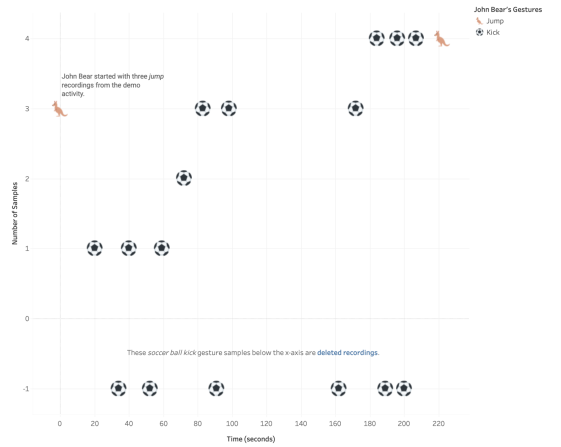
Another participant, John Bear, exhibited a very different approach by repeatedly rerecorded samples to make them as similar as possible. They created *jump* and then added *kick*. They quickly found that these two gestures interfered with one another. In John Bear's Gesture Log, notice that they removed and added *kick* six times. The gesture was interfering with *jump* due to similarities between the gestures' accelerometer plots.

Different Gestures that Look Similar

When Brian Bear described their process for recording three different sleep samples, they explained that it was based on how they sleep, which looks different throughout the night.



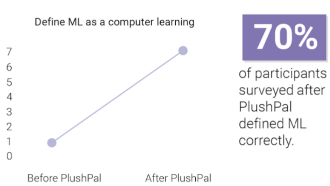
When this occurred, their initial response was to add more samples for *kick* and *jump*. This revealed that they were aware that the number of samples affects the model. After recording a new *jump* sample, they inspected the existing samples, counted the three peaks, and then proceeded to record another sample, counting aloud as they moved the bear up and down three times. This again shows their care to make the samples similar while interpreting the sensor data to match peak count with the number of jumps.



John Bear's case study featured an illustrative representation of the accelerometer charts and a gesture log.

Findings

Children were asked pre- and post-tests around topics related to machine learning. After their experience with PlushPal, 70% of children who answered the post-survey questions could provide a technical definition of machine learning (one that defines machine learning as a human training a computer to detect something specified). Before PlushPal, only 10% of children were able to provide such a definition.



Children and Data Literacy

Researchers saw through this study that children have different approaches to collecting data samples and highlight some misconceptions, including how machine learning models work. Children had a wide variety of gesture ideas to record, but could use support around how the quantity of sample sizes and similiarities between sample recordings impacts the accuracy of their machine learning models.

Findings

The shared results of the research project are summarized in text and a final chart made in Illustrator represents one of the key takeaways from the study.

Data

The data used in the visualization was collected from the study itself and fell into a few different categories:

- **Demographics**: The ages of participants, their previous programming experience (mostly with block-based editors like [Scratch]() and [MakeCode]()), and their previous experience working with a micro:bit.
- **Custom Gestures**: All 42 gestures and the stuffed animal projects that used them.
- **Workshop Data:** Adding and removing gesture samples throughout the duration of the workshop.

I had access to the data in the form of Google Sheets. I cleaned up the data there before exporting it as either a .xlsx files or .csv file, depending on the tool. One example of data clean-up was combining gesture names given by children to consolidate the list into 42 gestures. For instance, *run* and *running* were gestures on two separate projects that ultimately became the gesture *run*.

Once in the Observable notebook, I created Javascipt functions to manipulate the data further and match the links and nodes visualization style.

Tools

Charts were made in using D3 and Javascript in Observable notebooks, in Tableau, and in Illustrator. For software created, see Appendix B. Google Sheets was used for managing the data. The webpage is built in GitHub Pages using a Jekyll template and Markdown. Gifs were made using GIPHY Capture, and illustrations were made in Procreate.

User Testing and Results

Overview

The PlushPal visualization was tested with three prospective users over Zoom. The intention behind the tests was to identify what worked well with the design at the time of testing and what was challenging for users. Pre- and post-surveys were given to participants to gauge how the visualization changed their understanding of topics related to PlushPal as a result of seeing the visualization. Users were also given tasks to perform while looking at the visualization to determine if uncovering the answers was easy to do with the current design. Results from user testing informed changes made to the design for the final iteration of the interface.

Method

Participants were asked to sign a consent form and answer background questions to assess their prior knowledge on the subjects of machine learning, working with children, and hands-on learning. They were then given a pre-survey to complete via Google Forms. They read through the web page, and after they finished, they were asked to complete three tasks and were timed while doing so. A post-survey and qualitative questions with oral responses finished the user test. See Appendix A for a complete list of questions.

Results Analysis

Pre- and post-survey results indicate that the intervention of PlushPal improved participant understanding of content related to the study. The first question had no change, which may have been different results if the user testing included participants who were less familiar with children using technology. The second and third questions saw marked improvement between both surveys.

Question	Pre-Survey	Post-Survey	Change
S1. Do you believe children, ages 8-14 years old, are capable of understanding ideas around machine learning?	1.0	1.0	0
S2. After engaging in an experience with machine learning, approximately how many kids (ages 8-14 years old) do you believe could produce a technical definition of machine learning?	0.33	1.0	+0.66
S3: What tools or teaching techniques are needed to teach kids (ages 8-14 years old) machine learning and data literacy?	0.33	0.66	+0.33

For the third survey question, I coded the responses to determine whether an answer was sufficient. Phrases like ‘hands-on’ and ‘interactive’ felt reasonable, but ‘explanation’ alone did not.

Participant	Pre-Survey	Post-Survey
P1	Explanation	Interactive experience using an object that’s important to them
P2	Explanation, a computer	Explanation, micro:bit, computer
P3	Hands-on experience, access to resources	Hands-on experience with additional support

Tasks: Time Elapsed

All participants correctly answered the task questions, though their response times varied. The first task received the fastest response time on average. Tasks 2 (D3 visualization) and 3 (Tableau chart) had longer response times, which led to thinking around how to make these visualizations easier to understand and interpret.

Tasks	P1	P2	P3	Ave
T1. How many children who participated in the study had prior programming experience?	12 seconds	6 seconds	2 seconds	6 seconds
T2. What gesture or gestures were the most common between projects?	2 seconds	1 min and 16 seconds	13 seconds	30 seconds
T3. Describe Brian Bear's process while working. What was unique about their experience?	20 seconds	42 seconds	14 seconds	25 seconds

Qualitative Responses

What are aspects of this visualization that you appreciated?

Features that participants liked about the design include how the information was laid out and the overall look of the presentation. Details like the statistics and the graphical representations of the gesture data were compelling to see presented. Some of the participants liked the links and nodes interactive visualization, claiming that it was easy to navigate and gain insightful information.

What are aspects of this visualization that you found challenging?

One participant found the links and nodes visualization challenging and wasn't sure what to glean from the interaction at first. Two participants were challenged by the presentation of the accelerometer data, specifically Brian Bear's charts that showed three different samples for sleeping. They got lost in the shape of the graphs and didn't understand that they were each a separate sample. Some commented on labeling and titles for charts, which were small improvements made between user tests. One participant questioned the use of the gesture logs and didn't understand why some data points were below the x-axis (the deleted samples).

In your opinion, what's the big idea behind this visualization?

All participants' answers were similar to the idea of communicating about a research study that presented a tool for learners to engage with machine learning through a hands-on experience.

Discussion

The long response times for T2 and T3 and subsequent conversations with users around challenges in the design indicated that these two charts could use iteration. The D3 chart of the gesture linking was novel and clever to some, but mysterious and confusing to others. I reworked the prelude to this visualization in my final iteration of the design to clarify what the symbols and colors represented, and the significance of the interaction. The Tableau gesture log charts for Brian Bear and John Bear presented challenges to some users, so it was important that I revise these as well.

Revisions

I received feedback from Marti to take another pass at the accelerometer plots and gesture logs before the final submission. For Brian Bear's case study, I removed the accelerometer plot and replaced it with drawings to better communicate that the child recorded the gesture *sleep* in three different ways. I reworked the gesture log through additional text and introduced icons to represent gestures to communicate what the child's process was like when recording these gestures.

For John Bear, I simplified the accelerometer plot by making each axis a different shade of purple (users found the three different colors confusing). Paired the accelerometer plots for both gestures with illustrations to show that while the child saw two different gestures, the computer interpreted the data almost identically. I gave a similar treatment to John Bear's accelerometer chart as I did for Brian Bear, along with additional annotations on the chart itself to clarify a few data points (ex: why *jump* starts at three sample recordings and why *kick* has negative values).

Links

[PlushPal visualization]

[GitHub repository of the source code]

Attribution

I am the only person working on this assignment and completed all parts by myself.

Appendix

Appendix A: User Testing Questions

Participant Background

1. How old are you?
2. What is your prior experience working with children?
3. What is your prior experience with teaching or learning about machine learning?
4. What is your prior experience with hands-on or project-based learning?

Pre- and Post-Survey Questions

S1. Do you believe children, ages 8-14 years old, are capable of understanding ideas around machine learning?

- Yes
- No
- Maybe

S2. After engaging in an experience with machine learning, approximately how many kids (ages 8-14 years old) do you believe could produce a technical definition of machine learning?

- None
- One-quarter
- Half
- Three-quarters
- All

S3. What tools or teaching techniques are needed to teach kids (ages 8-14 years old) machine learning and data literacy?

Many options are acceptable, but answers that include direct experiences and working with materials align with the study. Insufficient answers are ones that limit direct instruction from a teacher, or no tools or techniques if the argument was that children are not capable of learning ML.

Tasks

T1. How many children who participated in the study had prior programming experience?
T2. What gesture or gestures were the most common between projects?
T3. Describe Brian Bear's process while working. What was unique about their experience?

Qualitative Questions

Q1. What are aspects of this visualization that you appreciated?
Q2. What are aspects of this visualization that you found challenging?
Q3. In your opinion, what's the big idea behind this visualization?

Software Created

Links and Nodes and Data Manipulation (Observable notebook)
Demographics Bar Chart and Isotypes (Observable notebook)