Design Jam

Date: Feb 23, 2024 2:00 p.m.
In attendance:

- Aesha
- Trang
- Quinn
- Tony
- Jesse
- Matt

With regrets:

- Mitch
- Zander
- Joe

Intent and Goals:

As we are starting to incorporate more things into the code base, some things can get lost in the cracks. The purpose of today is to make sure everyone is caught up on what we already have and align our goals in the second half of the term.

Agenda:

Where we are now:

Project as a whole and designs:

- Plans to go forward with a data abstraction layer for additional security between UI components and the IndexedDB.
- Lots of separate working components with no connection between them.
- Diagrams to be redesigned at a later date when ideas are more concrete.
- Matt went over project designs and project structure

Dev

- Currently on the web app, axis is displayed, data points can be hovered over, displays a label which will be filled in with entry fields for each point
- CSV data successfully read and accessed in IndexDB
- Interfaces (ex. DataPointProps) should be in their own file
- Jesse gave overview of currently deployed code
- Tony gave PCA overview

Build

- Working on ESLint
- Remove Node16? Not this sprint
- Dependency review
- Automated unit tests
- Deployment previews on PR (you can check what you're deploying)
- Deploy to Prod on Master (broken)
- Trang gave run-down of current build processes

Test

- Gherkins are used extensively to script integration tests
- Struggling because of the poor testing support for VR
- Addressing Drei/Jest conflict
- Extended Typescript's assert function
    - The purpose of asserts is to stop a program if the logic of events is incorrect
    - Typescript implements asserts as an interface for you to implement your own logic as
    - In order to allow the ability to disable at runtime for productions sake, must implement flags
    - We do have it implement currently
- Logging configs aren't loaded with Vite

Where we are going:

Data abstraction layer designs:

- Transpose the incoming data to allow greater control over columns/fields selected, hopefully reducing load/read times.
-

https://app.diagrams.net/#G1pue7o1hbV5-o876cVSWB4kSA5G7GOVA3#%7B%22pageId%22%3A%22CqYIRr6Jr4gl-h0RM-nM%22%7D

Test

- Looking into manual plus auto testing for in VR mode
- Proper smoke and sanity testing

Notes:

- Questions to ask Dr. Osgood:
    - What specific characteristics need to be displayed for each data point?
    - Do we show values of all fields or only the 3 selected fields?
    - Do we show all PCA components with their values?
    - Do we need to show the info like colour/label of the points as well?
    - What to do if there are any missing data points or outliers in the dataset?
        - This is a Matt answer. Right now we are dropping values, but eventually they asked us to have a special spot on the axis for them and have the ability to show/hide them.
    - The example of PCA in the spec looks like linear regression. What are we missing?
      → e.g., 0.1*age+0.2*income-0.6*income + 5.0*weightInKg

Email sent:

1. What specific characteristics need to be displayed for each data point? Is it to display all associated field values for the point, or the values of the fields currently being plotted?
2. Do we show all PCA components with their values?
3. Do we need to show the info like colour/label of the points as well?
4. We had some confusion that came up in our PCA implementation. The example of PCA in the spec looks like linear regression. (→ e.g., 0.1*age+0.2*income-0.6*income + 5.0*weightInKg) We were wondering where the coefficients of this equation come from, as is doesn't really match any descriptions of PCA that we could find, as they dealt with covariance matrices. What are we missing?

And here are the answers I received:

1. There could potentially be many fields which would lead to a cluttered view if all were displayed. The spec says, “enumerate information on the points for fields selectable for this purpose.” This suggests to me that a subset of fields is appropriate. I am going to suggest that the subset include the fields bound to {x, y, z, shape, colour} plus a small (e.g., 1–5) number of other fields that the user selects.
2. No. There could be too many to feasibly display them all simultaneously. Once the PCA has been undertaken, the principal components should be available as additional “fields” that can be associated with an axis, colour, shape, or label.
3. Yes.
4. Please ignore that example in the spec. The output of a PCA is a series of unit weight vectors that map the original coordinates to the new coordinates, so those coefficients should sum to 1. If original coordinates are (x,y,z) and scores of the first principal component are t = (t1,t2,t3) then w = (w1,w2,w3) such that t1 = w1*

x + w2 * y + w3 * z. In this way it may seem similar to a linear regression because it is a linear combination. Not sure if this is helpful, perhaps we can discuss further on Thursday.

Actionable Follow Ups

Item	Person
Flush out data abstraction	Matt
Set up another sys arch meeting	Matt
Remove assertions from try-catch	Tony
Get test update on Drei/Jest issue	Mitchell
Compile the rest of the questions for the next stakeholder meeting/email	Aesha
UI updates (maybe look at 3D space stuff???)	
Follow up that IndexedDB is supported on Oculus browser	Aesha/Jesse/Quinn
Follow up on Oculus filesystem CSV import	Quinn
CSV file loader can’t get data into IndexedDB when url data is loaded	Tony