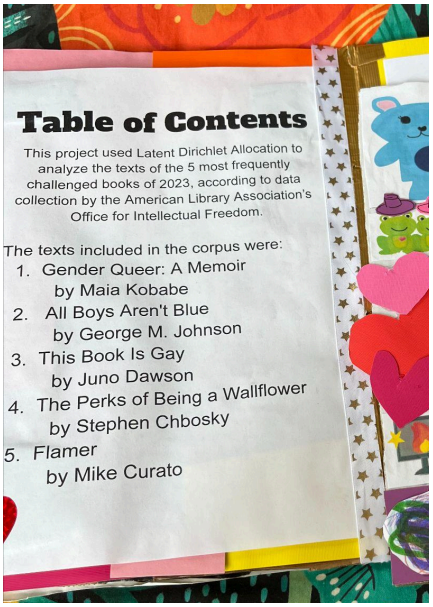
Data Physicalization - This book is Challeng(ed/ing) Final Reflection
Maddy Casey

Encoding Data as Images

While preparing the texts for use was the most time consuming component of this project, selecting images to use for the encoding of each word proved to be the most challenging component. Specifically, the reality that a large number of words which appeared in the model were verbs or else referred to an abstract concept such as “sense” in another way made encoding them as a literal image somewhat difficult. This challenge is a product of working with words or other qualitative ideas as data- the encoding process may not be precise- rather, much like the music encoding project I completed with my family’s mood data, subjectivity and personal choice of encoding form an essential part of the process. Ultimately, during this class I learned that the way in which we choose to encode data as designers or data scientists is always a matter of choice. In this project, the subjectivity of that decision making process was increasingly apparent to me during the creation process- this reality leads me to believe that similar work might be a good teaching opportunity to help students develop Creative Data Literacy skills (in the thread of Catherine D’Ignazio’s work), given that the assignment of data points to an image is so clearly subjective, and thus a good opportunity for illustrating the choices that go into all data representation.



Use of Topic Modeling

Topic modeling is a tool which can illustrate patterns in word choice across a selection of texts, illuminating shared meaning. To my mind, the project and model effectively did so; for example the fact that the most challenged books share terms related to sexuality and relationships, is a valuable insight. However, by focusing on the terms and ideas that the books have in common, this practice of analysis also eclipses some of the distinctions between the texts- the nuances that make them uniquely valuable as literature, as stories or guides written for marginalized young people. Additionally, the project’s process also encouraged me to reflect upon the use of a digital tool for analysis as a stepping stone for the final project. Specifically, one of the

affordances of data physicalization that I discovered over the course of the summer, is the reality that the work of making data machine readable is not always necessary for the creation of powerful materializations of structured information; in fact, this dimension of the work can make data physicalization far more accessible as a practice to people who do not have as much experience with data analysis tools and work. Thus in future work, I would consider alternate ways of engaging with this data in a way which could effectively point to the subjectivities of reading a text. For example, a crowdsourced version of this project could be developed, wherein participants are asked to select or create images representing portions of a text (ie., participants are assigned to read a chapter and select or create images of the significant words or concepts they found in the text). While this version of the project would be potentially difficult to plan, and more subjective, in my mind, the subjectivity inherent in the ways in which people engage with texts through the practice of reading, is an essential component of understanding meaning-making in literature.

Opportunities for Future Iterations

In future iterations of the work, I would add additional collages, representing more topics. For example, another interesting topic I would like to represent is “Like, Time, Having, Friends, Coming, Community, Away, Used, Way, Able, Taught, Came, Students, Best, Teachers”- Community Support. I also realized during the artist talk that I needed to think through the way in which I spoke about the work that I conducted and the way in which I explained the process in the table of contents section using more accessible language to the audience. While I chose to use some more jargon (including “topic modeling” and “corpus”) during the presentation to a group of information scientists, following the presentation, I realized that not everyone in the audience was familiar with these concepts, or that at the very least, their use might make the project a bit inaccessible or illegible. Thus in future, I would explain the process of the project in the “Table of Contents” more simply, “ie., this project used topic modeling, a form of analysis which detects patterns in word-use across a series of texts, to demonstrate trends and shared concepts in the most challenged books from 2023”. In future iterations of the project, I would also like to ensure that there is a little bit less “noise” on the project; specifically, I would like to devote some time to thinking through how to make the design on the cover of the book, and the book’s spine into data visualizations in their own rights.

At the same time, however, I do think that the “sex/relationships" collage and the cover were aesthetically interesting, if not perfectly polished. Aesthetics here served as an important component of a successful physicalization. The choice in the design of the cover, which is somewhat evocative of the “burn book” from *Mean Girls* (2004) and the choice to use collage as a methodology, which is somewhat evocative of Zines, also intentionally evoke an ethos of creation and personal expression which is rooted in youth and queer culture, matching the subject matter of the most challenged books/dataset. I also intentionally used color and texture in the “sex/relationships” collage - using pinks, purples, and reds, and repeated heart shapes to emphasize the topic in question. In future collages, I would like to continue to take time to think through use of color, shape and texture to effectively highlight the overall concept in each collage.