Process Documentation

Unravelling the thread: Understanding the rhetoric between clothing and sexual assault in India, through the eyes of the law and survivor .

Julius Stein and Priyanka Pillai

Master in Design Engineering - Collaborative Design Studio Oct 12th, 2022

Index

01 Project Background

02 Within Clothing

03 Design Values and Objectives

04 Design Process

05 Technical overview

06 Visualisation experiments

07 Mood board

08 Ideation

09 Important findings and insights

10 Making: Key Tools

11 Making: Key Methods(Data Gathering & Cleaning)

12 Making: Key Methods (Data Processing)

13 Making Final visuals - few examples

14 Making - example visual

15 Final screen examples

16 Future Impact

17 Reflection

Project Background

The discourse around sexual assault, sexuality, and sex positivity is often centered around consent and male entitlement. A prevalent belief (and legal defense) in this gendered society is the prevalent notion that the choice of clothing often contributes to instances of sexual assault and rape. More specifically, attire has been frequently cited in legal cases as a form of “implied consent”

As part of our project, we aim to collect and analyze data, both quantitative and qualitative, to determine the validity and indeed the accuracy of this notion. We hope to show that this rhetoric between what one wears and consent is grossly misunderstood and that, moreover, it is compounded upon by institutional structures such as dress codes and legal privileges. Through a series of interactive visualizations, we intend to illustrate the relationship between self-image and external appraisal of one’s clothing, and the way in which these perceptions are codified.

Our research methods will include a mixed methodology of literature review of published papers and articles, police reports, the study of legal cases, along with stakeholder interviews. We will additionally conduct a large-scale study using Amazon Mechanical Turk to gauge perceptions of clothing articles by gender, assess underlying biases, and create a central dataset that can attempt to simultaneously address the various lenses with which prior studies have approached this topic. We will finally visualize the relationship of our study to prior work in the field in order to suggest a causal link.

Within Clothing

The scale and pattern

2021 Over the past year in India, **rape in India rose by almost 20% - averaging around 87 cases per day**. Scores of additional cases go unreported each day. Gendered violence In the same timeframe, crimes against women rose by 63%. These numbers reflect a rising trend of gendered violence throughout the country. Finally, the poor conviction rates, victim-shaming practices. , and persisting misogyny point to a legal system that enables gender-based violence, and on occasion, even endorses it.

Gap and Opportunity area

One important facet of this discourse is clothing, which is a manifestation of choice, identity, and freedom.

Clothing as a defense - Little work, if any has investigated how clothing affects legal outcomes Research has established the ways in which we hold differing conceptions of ourselves and of others depending on one's dress.

The opportunity

While personal narratives have been ubiquitous in this discourse, there have been few narratives that have combined the larger data to the personal narratives to drive home the need for serious interventions.

Value Systems

01 `Sensitivity is key`
In dealing with and representing information, especially personal narratives.

02 `Strive for multiplicity`
To capture as diverse and nuanced experiences as possible, within different contexts and at varying scales.

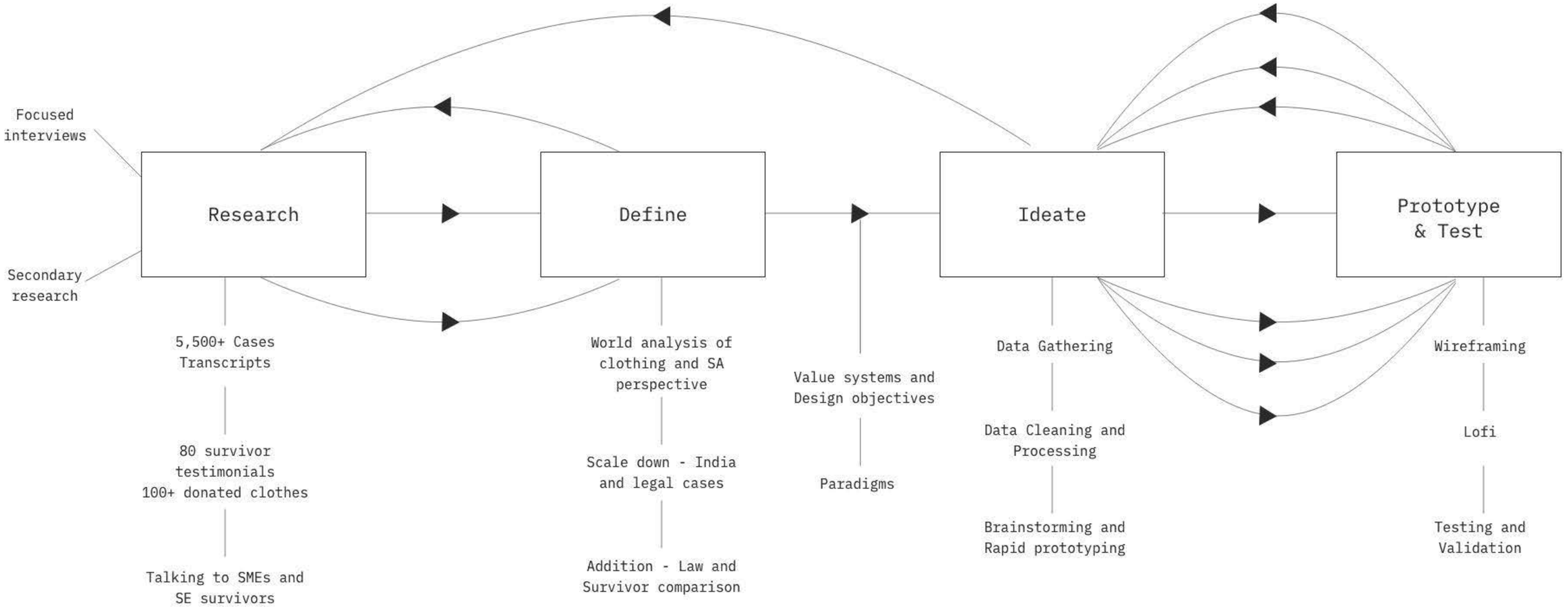
Design objectives

01 Reflect the **scale** at which sexual assault is validated with victim-blaming (especially clothing and sexuality), across institutions.

02 Showcase **distinct perceptions**, while also highlighting differences between perceptions and lived experiences.

03 Identify **language and verbiage** that emerge when sexual violence is systemically enabled.

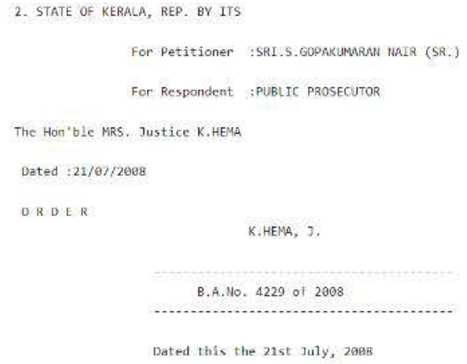
Design Process



Technical overview

When we began our research, we started by looking at the existing work which addressed the way in which attire directly relates to sexual assault. Researchers had thoroughly examined the relationship between both the conception we have of another when they wear different clothing, as well as the effect on self-image that dress has. Yet little research had been conducted on the way in which the victim's dress is systemically institutionalized either through dress codes or clothing bans, or in court cases where the question 'what was she wearing' is subversively used to introduce doubt or cast blame on the victims. We initially intended to examine the United States, where we found by far the largest volume of case transcripts and case data, but although the US government's Public Access To Court Electronic Records (PACER) seemed to be a promising start, searching for a name or term costs "$30.00 per name or item searched plus $0.10 per page per document delivered electronically", making even small-scale analysis of thousands of cases relatively unaffordable to the public who it claims to provide access to. As a result, we turned to India, where a far larger population provided ample cases to analyze for relatively little cost through a series of automated pipelines.

Our technical workflow is discussed in further detail, but at the largest level, our data came as a list of unstructured text bodies from IndiaKanoon, our processing, documentation, and workflow is hosted on GitHub (https://github.com/JuliusStein/kanoon_transcript_nlp), and our main dependency for analysis was GPT-3.


2. STATE OF KERALA, REP. BY ITS

For Petitioner :SRI.S.GOPAKUMARAN NAIR (SR.)

For Respondent :PUBLIC PROSECUTOR

The Hon'ble MRS. Justice K.HEMA

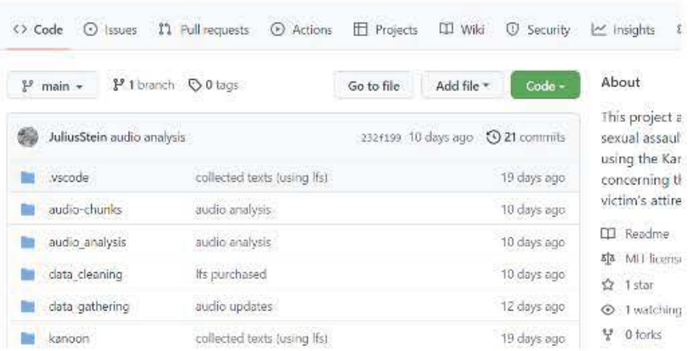
Dated :21/07/2008

O R D E R

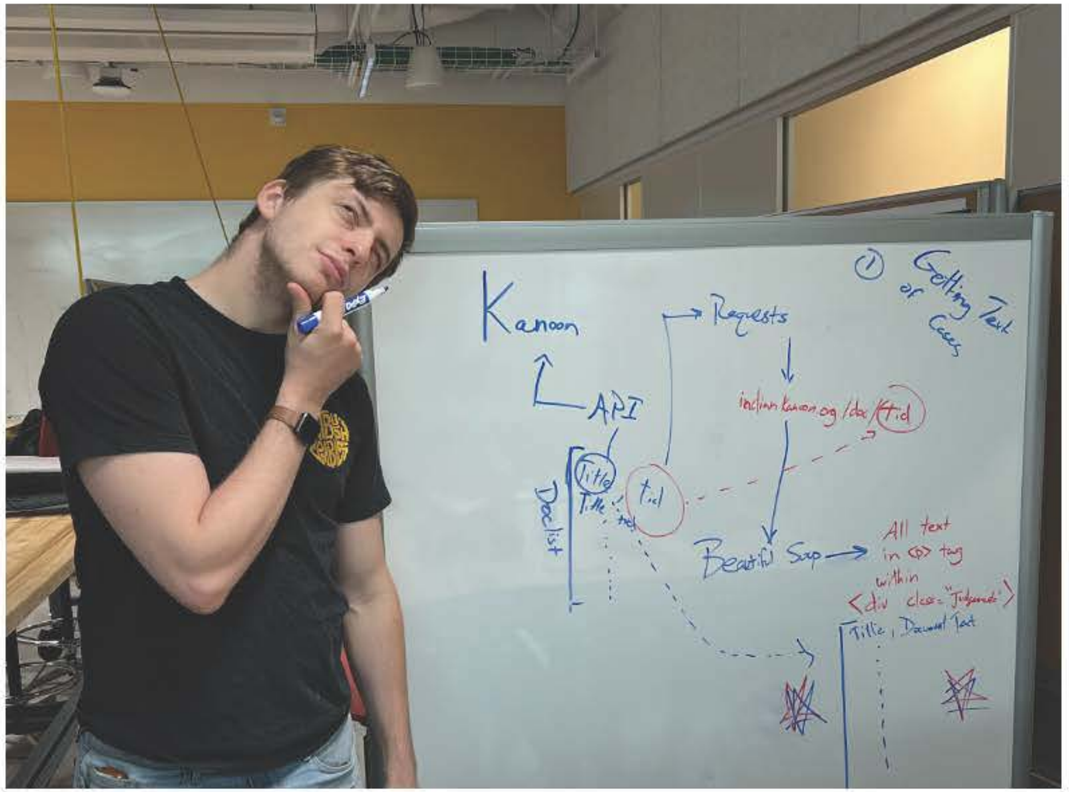
K.HEMA, J.

B.A.No. 4229 of 2008

Dated this the 21st July, 2008

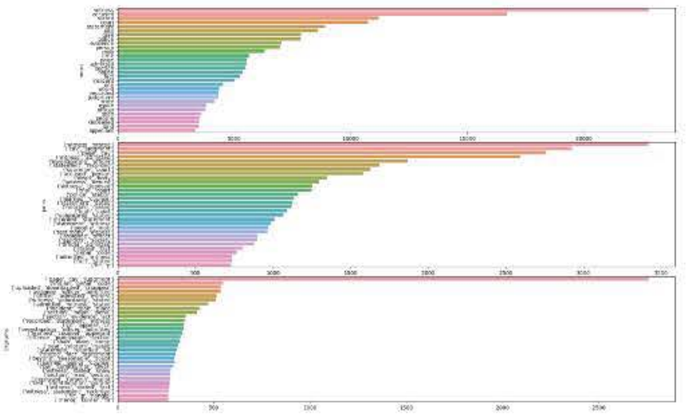


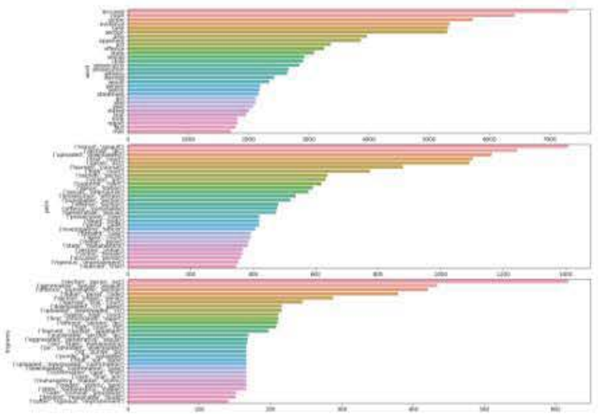
Technical overview

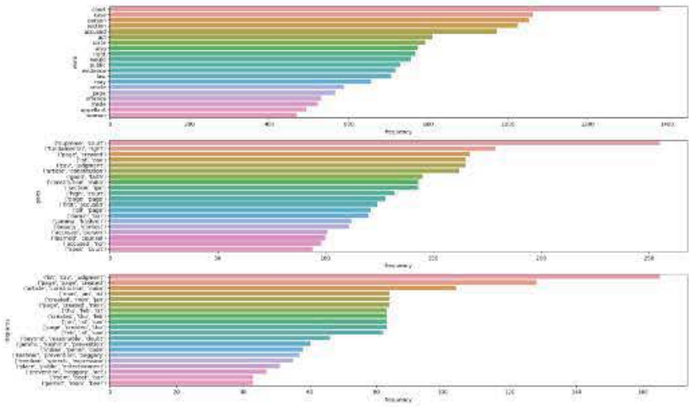


Visualisation experiments

```
def __init__(self):
    self.base_url = "https://api.indiankanoon.org/"
    self.auth_token = "0ea7d17df82f0c9206f58113b395f3f28484334a"

    self.headers = {
        'authorization': "Token {}".format(self.auth_token),
        'cache-control': "no-cache",
    }
    self.api_session = requests.Session()
    self.api_session.headers = self.headers

def search(self, formInput, pagenum=0,
           fromdate=None, todate=None,
           title=None, author=None,
           cite=None, bench=None):
    #  Creating parameters
    params = {
        'formInput': formInput,
        'pagenum': pagenum
    }
    if fromdate:
        assert isinstance(fromdate, datetime)
        params['fromdate'] = fromdate.strftime('%d-%m-&Y')

    if todate:
        assert isinstance(todate, datetime)
        params['todate'] = todate.strftime('%d-%m-&Y')

    # Making the request
    response = self.api_session.post(
        urljoin(self.base_url, 'search/'), params=params)
    response.raise_for_status()
    return response.json()
```







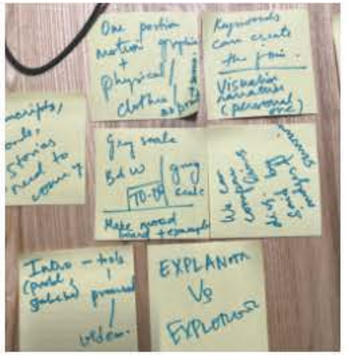
Important findings and insights

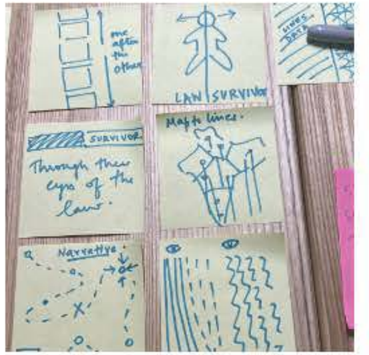
01 Our initial selection of this topic was to design and create for an interest area. As we dove deeper, we were overwhelmed at the sheer scale of this social issue. As it stands, this project is now an affirmative initiative that is trying to tackle gender-based violence - **it is a start.**

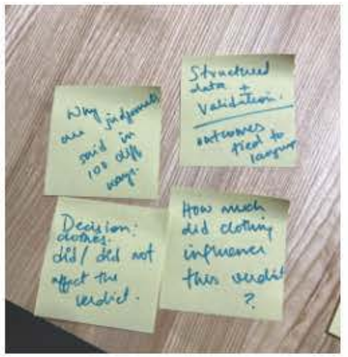
02 As we uncovered more cases where clothing was a prominent defence strategy, we came across the data set of acquittal rates. What was even more interesting was that **certain judges had higher acquittal rates for sexually violent crimes than others.** While this implied bias, we were unsure how to capture and represent this - for now, it is a potential area of expansion.

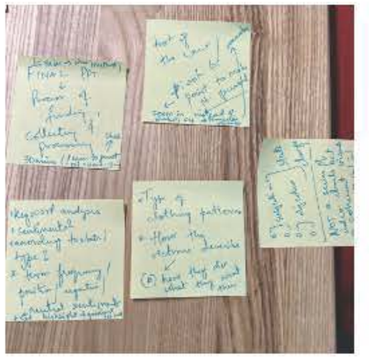
03 When we started our research, we knew that clothing was used as legal defence and social justification. But we were still unprepared to learn how this was also used to cast aspersions and malign the character of survivors. **This drove our decision to provide a platform for viewers to hear directly from the survivors themselves.**

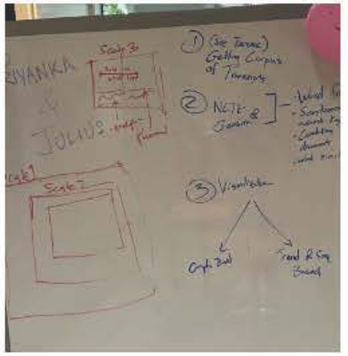
04 The survivor testimonials were incredibly hard and complex to work with, especially as a data set. **While the 'eyes of the law' section dealt with clothing and body, the survivors spoke mostly of their feelings and cognition - this was difficult to validate and quantify.**

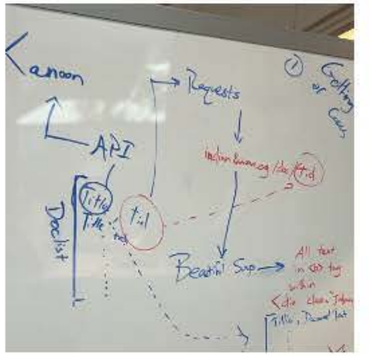




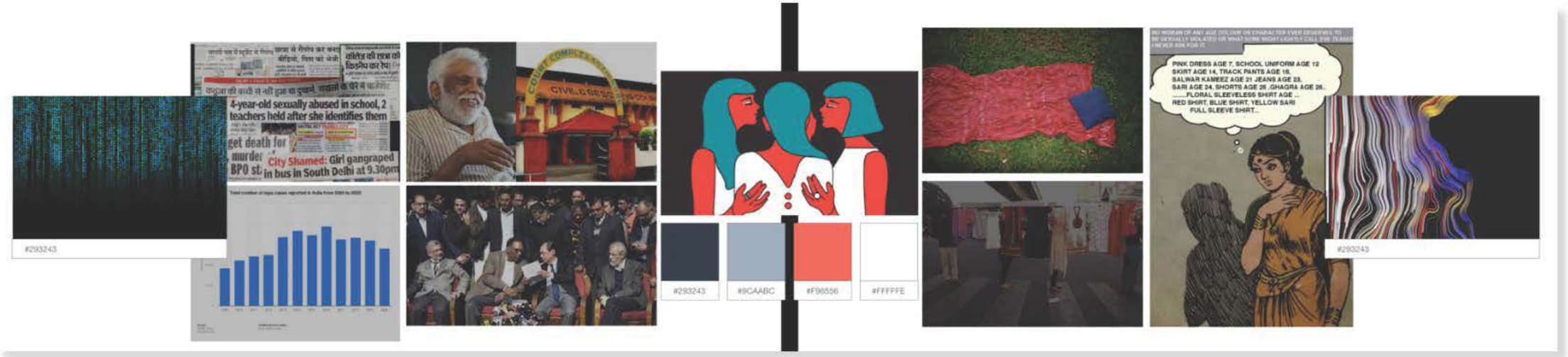




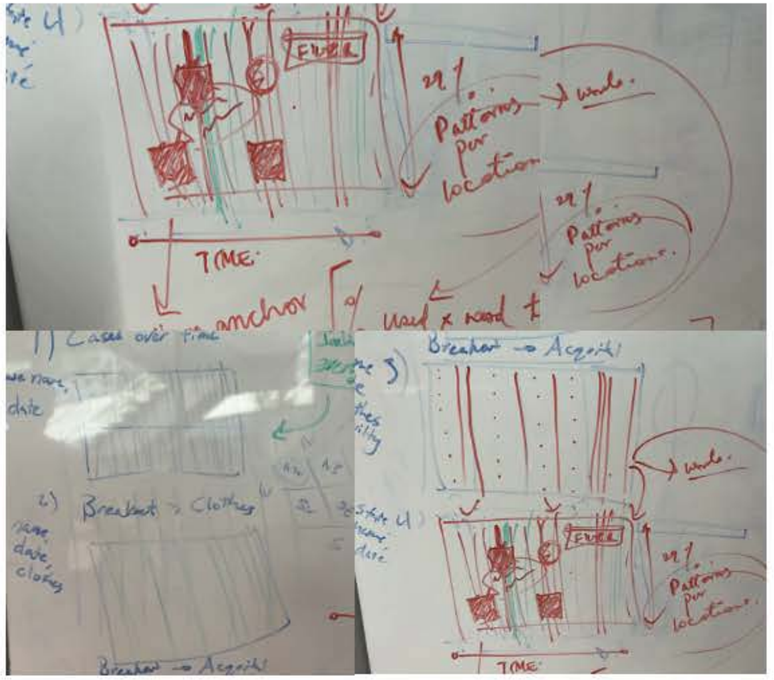


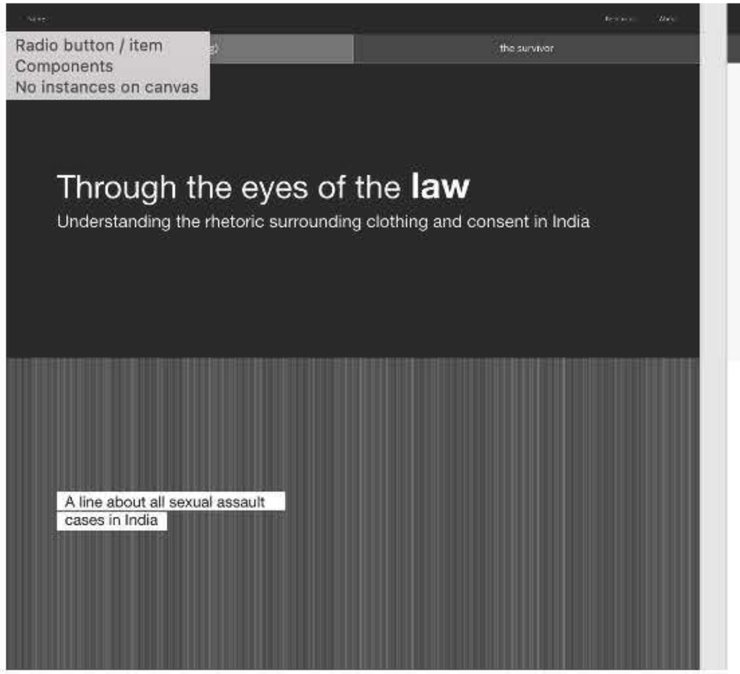


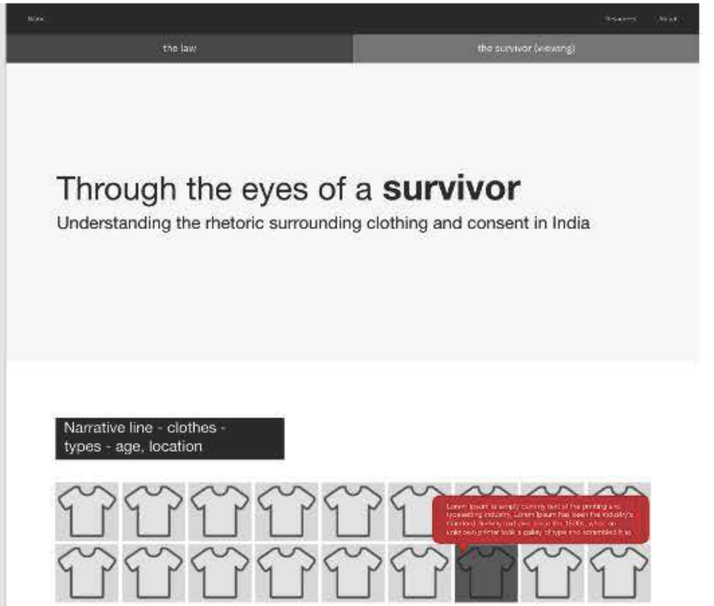
Mood board


#293243
4-year-old sexually abused in school, 2 teachers held after she identifies them
City Shamed: Girl gangraped in bus in South Delhi at 9.30pm
#293243
#8CAABC
#F98556
#FFFFFE
PINK DRESS AGE 7, SCHOOL UNIFORM AGE 12
SKIRT AGE 14, TRACK PANTS AGE 18,
SALWAR KAMEEZ AGE 21 JEANS AGE 23,
SARI AGE 24, SHORTS AGE 25 ,GHAGRA AGE 26..
.....FLORAL SLEEVELESS SHIRT AGE ...
RED SHIRT, BLUE SHIRT, YELLOW SARI
FULL SLEEVE SHIRT...
#293243

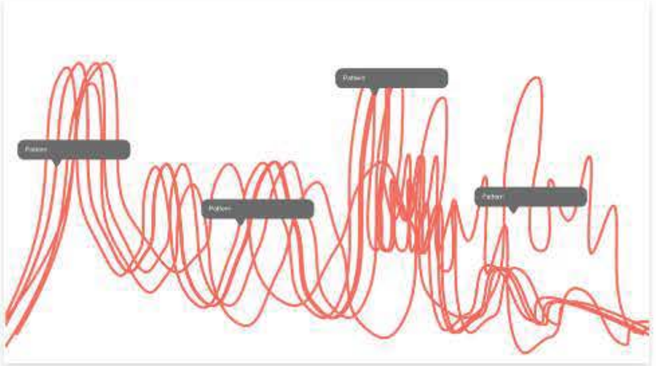
Ideation

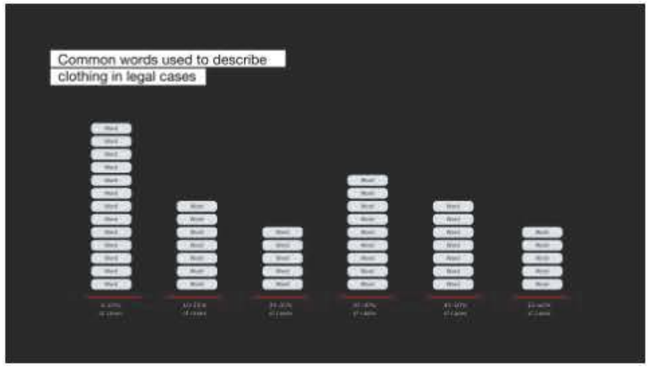

FIVER
TIME
Breakout → Acquit
Breakout → Clothes

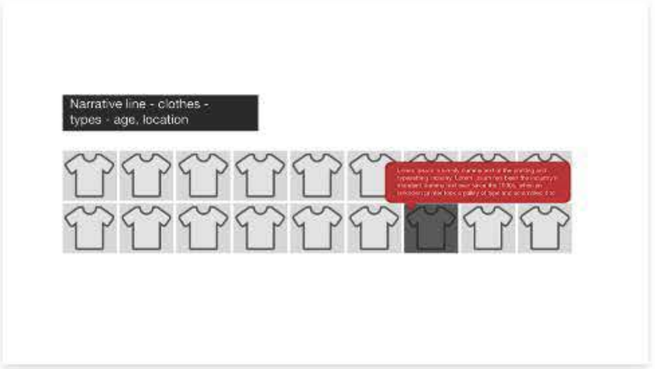

Radio button / item
Components
No instances on canvas
the survivor
Through the eyes of the **law**
Understanding the rhetoric surrounding clothing and consent in India
A line about all sexual assault cases in India

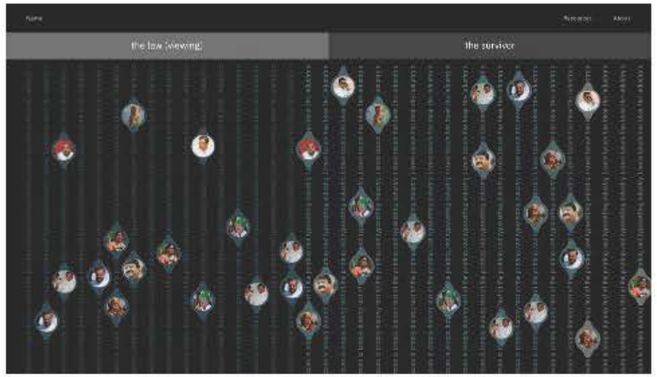

the law
the survivor (viewing)
Through the eyes of a **survivor**
Understanding the rhetoric surrounding clothing and consent in India
Narrative line - clothes - types - age, location

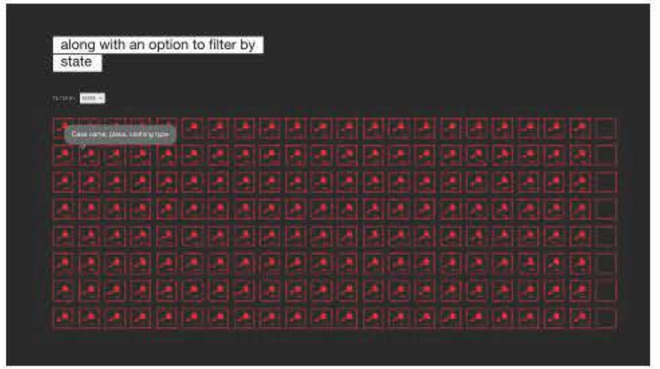
Ideation (cont)











Making: Key Tools

Technical overview

Data Collection: Indian Kanoon

01 Metadata providing the date, judge, and courthouse of each transcript. Analyzed all transcripts returned for the query "sexual assault" which returned roughly 5500 cases.

Processing Pipeline: OpenAI

01 GPT-3 to read and analyze the outcome of each case & label our cases according to the presence of clothing in the transcript.

Data creation

Transcript Analysis: NLTK

01 Analyzed their contents through a variety of natural language processing approaches to search for patterns in the language used in cases where clothing was mentioned.

Data for further analysis

01 Examined statistical correlation between mentions of clothing and acquittals, regional differences in case outcomes, acquittal frequency over time, etc.

Making: Key Methods (Data Gathering & Cleaning)

Much of our early work involved querying Kanoon and gathering the case data. Because our free license would only allow us to query a few hundred full texts in addition to returning the id's returned from an initial query, we set up a pipeline to make new accounts for Kanoon which would query a different subset of the pages of cases, storing each as in a Pandas dataframe and uploading them all joined as one JSON list using GitHub LFS. We then created an automated pipeline to label each case for the key missing parameter - the case outcome - as well as the presence of language around dress or clothing. After all 5500 cases were queried, processed, and labeled, we explored the trends that we could find from disaggregating the data by the date, the presence of clothing, the region or court, and the particular judge who tried each case. The analysis revealed a number of concerning discrepancies in conviction rates across each of the desegregations, though ultimately our initial hypothesis that cases which discussed the victims' attire would be less likely to lead to an acquittal were disproven.

	title	tid	court	date	judge	full_text	resolution	dressMentioned	resultCode	dressCode
0	Madan Gopal Kakkad vs Naval Dubey And Anr on 2...	1314858	Supreme Court of India	1992-04-29	Bench: Pandian, S.R. (J)	CRIMINAL APPELLATE JURISDICTION : Criminal App...	Conviction	No	C	N
1	Santhosh Madhavan @ Amrutha ... vs Circle Insp...	1815206	Kerala High Court	2008-07-21	unknown	A girl aged 14 years is the alleged victim in ...	Neither	Yes	U	Y
2	Atender Yadav vs State Govt Of Nct Of Delhi on...	151172763	Delhi High Court	2013-10-29	Author: Kailash Gambhir	KAILASH GAMBHIR, J. 1. By this appeal filed un...	acquittal	No	A	N
3	Beeru vs State Nct Of Delhi on 11 December, 2013	13922091	Delhi High Court	2013-12-11	Author: Kailash Gambhir	% KAILASH GAMBHIR, J. 1. One can visibly see t...	Acquittal	No	A	N

Making: Key Methods (Data Processing)

All of our data processing was done in python. Beginning with the dataframe shown previously, we analyzed the textual trends using a bag of words model from genism's implementation of Word2Vec, in which each token within the corpus is turned into a vector that can be compared to other vectors through cosine-similarity. Separately, we plotted the line-trends using matplotlib by simply positioning each case temporally and using the labels generated by GPT-3 to color and filter them. Audio was analyzed similarly, though we first needed to use googles text to speech to cut fragments of the audio, translate each to a string, and then join all strings after the full audio had been parsed.

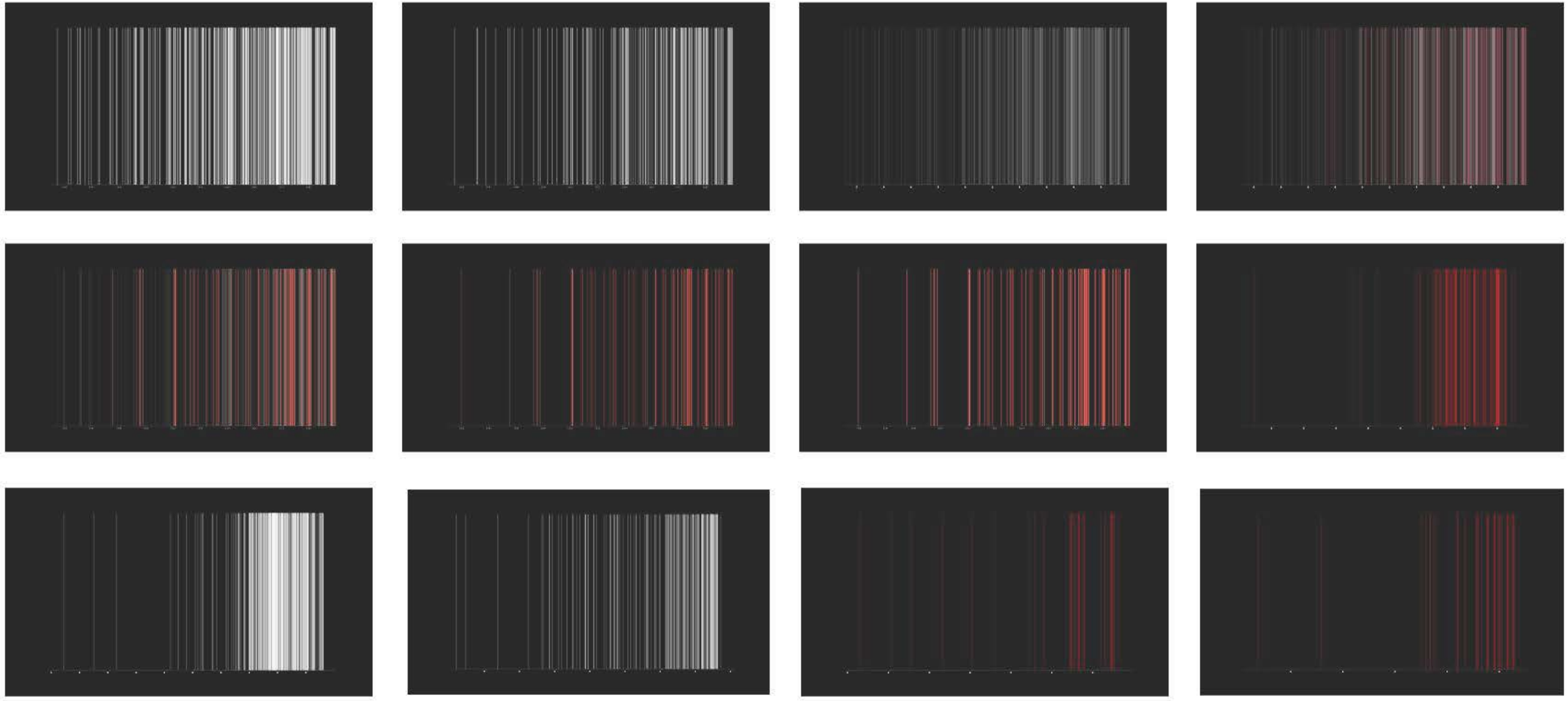
```
import gensim
from gensim.models import Word2Vec

f = alltexts.replace("\n", " ")
data = []
print("Tokenizing Sentences...")
sentence_tokens = sent_tokenize(f)
# iterate through each sentence in the file
for i, tok in enumerate(sentence_tokens):
    temp = []
    if(i%10000 == 0):
        print(i,"of",len(sentence_tokens), 'tokens complete')
    # tokenize the sentence into words
    for j in word_tokenize(tok):
        temp.append(j.lower())
    data.append(temp)

# Create Common Bag Of Words model
print("All tokens generated! Building Bag of Words model...")
model1 = gensim.models.Word2Vec(data, min_count = 1,
                                vector_size = 100, window = 5)
```

```
for i, audio_chunk in enumerate(chunks, start=1):
    # export audio chunk and save it in
    # the `folder_name` directory.
    chunk_filename = os.path.join(folder_name, f"chunk{i}.wav")
    audio_chunk.export(chunk_filename, format="wav")
    # recognize the chunk
    with sr.AudioFile(chunk_filename) as source:
        audio_listened = r.record(source)
        # try converting it to text
        try:
            text = r.recognize_google(audio_listened)
        except sr.UnknownValueError as e:
            print("Error:", str(e))
        else:
            text = f"{text.capitalize()}. "
            print(chunk_filename, ":", text)
            whole_text += text
```

```
for index, row in caseFrame.iterrows():
    date = row['date']
    year, month, day = map(int,date.split(" "))
    if(year>1970 and index%5==0):
        if(row['resultCode'] != "U" and row['dressCode'] == "V"):
            if(row['resultCode'] == "A"):
                try:
                    plt.axvline(dt.datetime(year, month, day), color="red",linewidth=0.5)
                except:
                    print("Y:",year,"M:",month,"D:",day,"failed")
            else:
                try:
                    plt.axvline(dt.datetime(year, month, day), color="white",linewidth=0.5)
                except:
                    print("Y:",year,"M:",month,"D:",day,"failed")
```

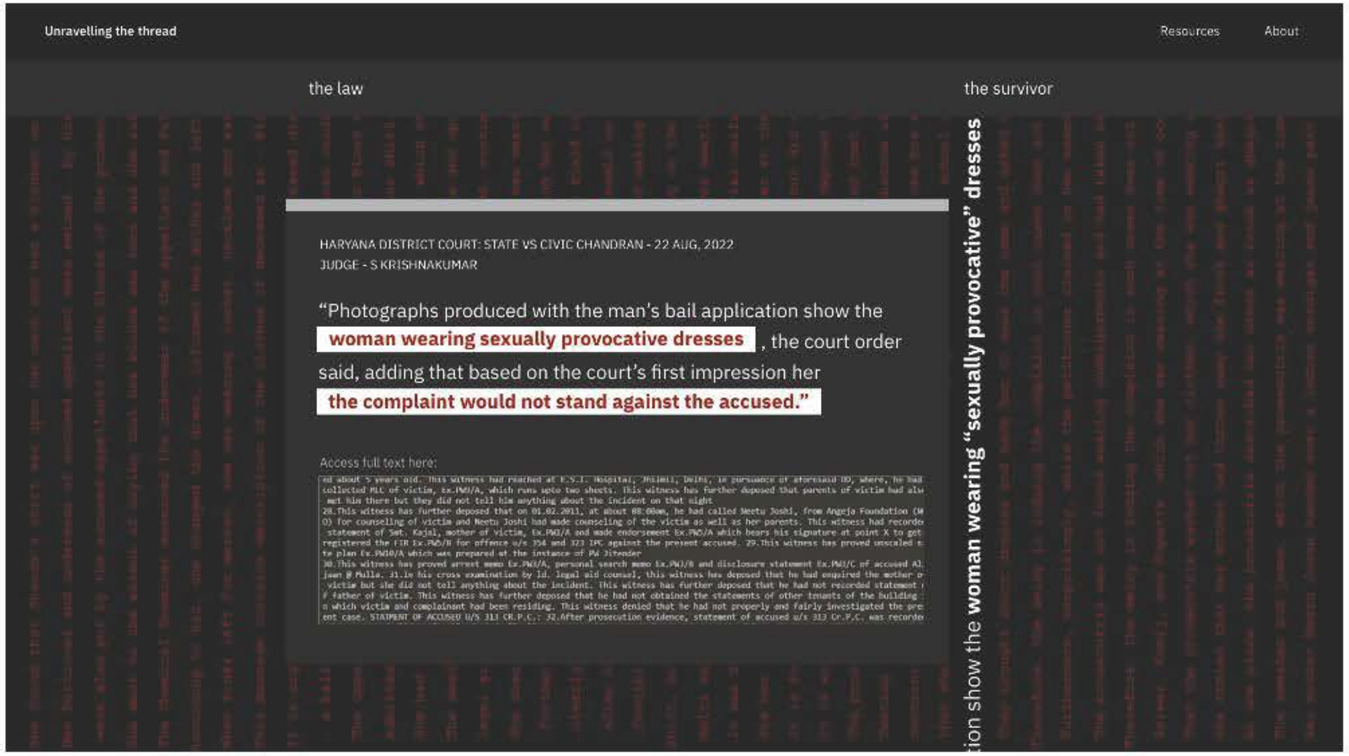
```
new_tokens = word_tokenize(sentence)
new_tokens = [t.lower() for t in new_tokens]
new_tokens =[t for t in new_tokens if t not in stopwords.words('english')]
new_tokens = [t for t in new_tokens if t.isalpha()]
lemmatizer = WordNetLemmatizer()
new_tokens =[lemmatizer.lemmatize(t) for t in new_tokens]
#counts the words, pairs and trigrams
counted = Counter(new_tokens)
counted_2= Counter(ngrams(new_tokens,2))
counted_3= Counter(ngrams(new_tokens,3))
```

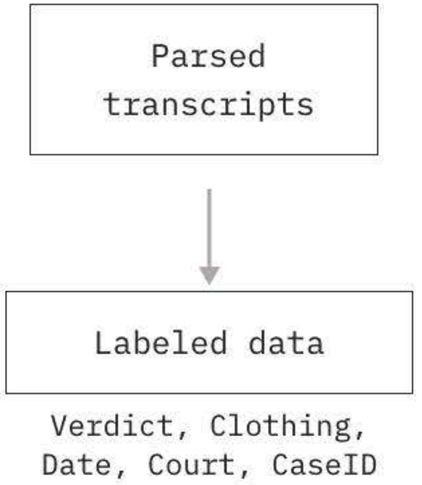
Making Final visuals - few examples



Making - example visual

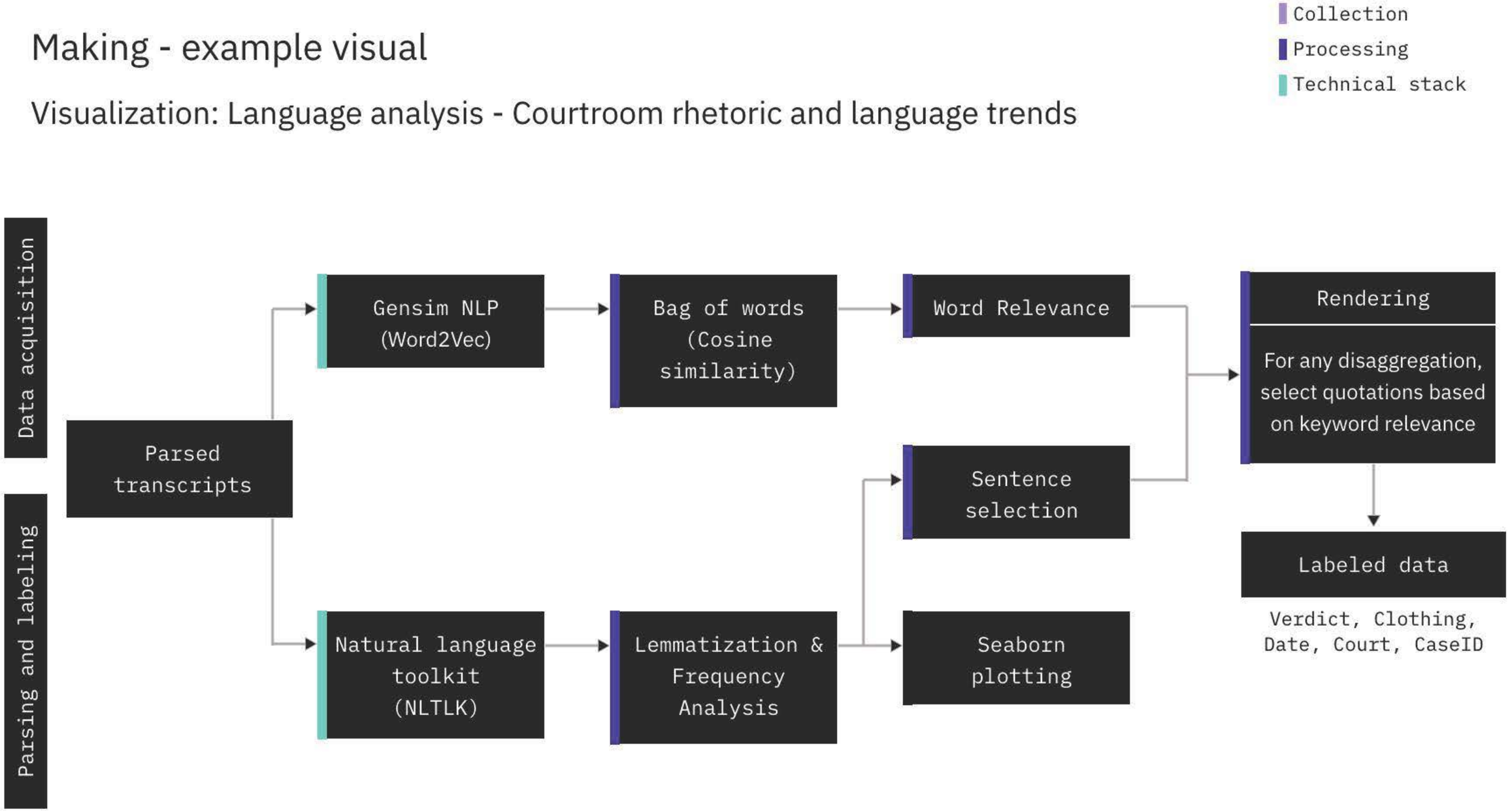
Visualization: Language analysis - Courtroom rhetoric and language trends





Making - example visual

Visualization: Language analysis - Courtroom rhetoric and language trends



Making - example visual

Visualization: Language analysis - Courtroom rhetoric and language trends

`Side by side comparison`

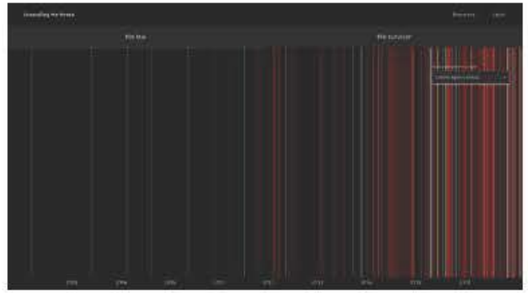
law (example)

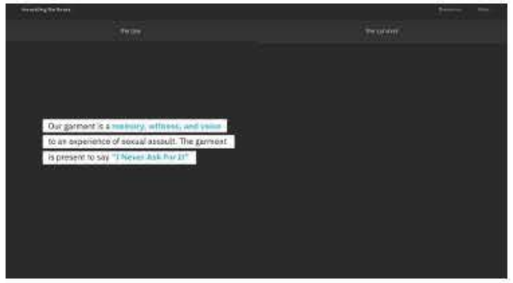
“Photographs produced with the man’s bail application show the **woman wearing sexually provocative dresses,** the court order said, adding that based on the court’s first impression her **the complaint would not stand against the accused.”**

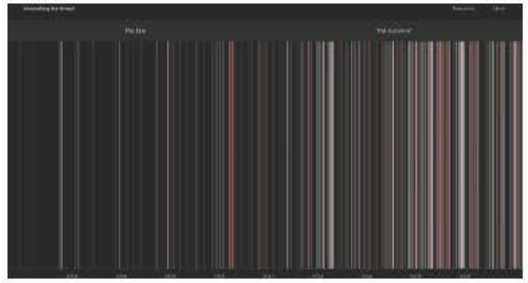
survivor (example)

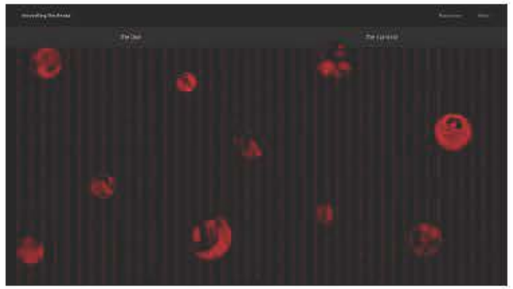
“What part of me is dressed in any way to give the idea that I’m available for that to happen me?”

Final screen examples





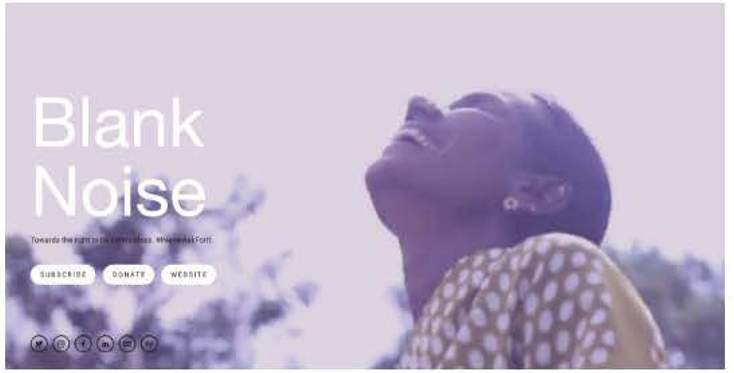


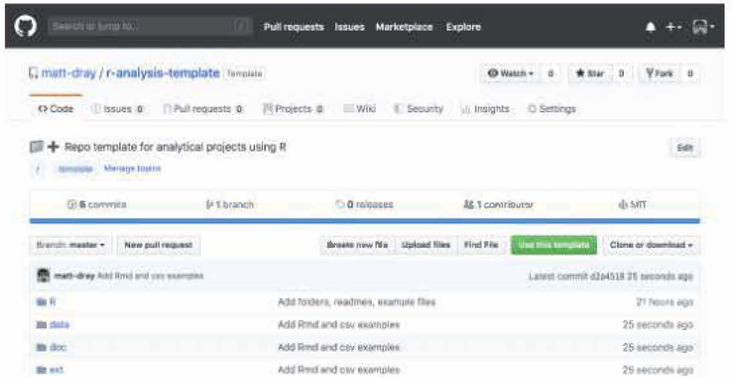


Future Impact

Collaboration with grassroots organisations - Blank Noise

Based out of Bangalore, India, Blank Noise a public art and social rights campaign that seeks to combat sexual harassment.
We aim to work with Blank Noise and other such organisation to collect more narratives, diversify our information, concrete with survivors, and make our project more intersectional.



Scale globally to introduce threads of race, culture, and colonisation

Gender-based violence is not disjunct from race and culture. As clothing has been an important export of colonisation, we seek to understand its implications on sexual violence. In this interest, we intend to host the labeled dataset that we used for others to do further analysis with.