

~~censored~~

<https://youtu.be/UGsyOfj2iNM>

<https://github.com/ashwinv11/censored>

Introduction

India's government is becoming open source and is releasing a lot of government data to the public through the Digital India Initiative (OGD). Developers can access this data through a series of APIs and CSV files from their online portals (OGD). One of the datasets that have been released, but haven't been touched by anyone, is about rape and sexual assault statistics in India (OGD). Some examples of the data that can be accessed are:

- 1) Number of rape victims per district, per state, per year, and per age category (OGD).
- 3) Whether the attacker was known to the victim, whether they were a family member/extended family member/neighbor (OGD).
- 4) Data on people arrested because of rape accusations (OGD).

According to the website's statistics, the ratio between the most popular dataset being accessed, and the average datasets pertaining to rape and/or sexual assault is close to 1200:1 (OGD). This is because the format in which the data is released is inaccessible and unusable to the general public. Users need to create an account to access the documents, and accessing and finding a specific dataset can take up to hours because of their poorly implemented search tool (OGD). Even the way the datasets are indexed, are confusing i.e. there are too many parameters to search through in order to find the information you are interested in (OGD). This

data is from 2001 to 2012 so even though it hasn't been updated in 3 or more years, at least it's from a credible source, i.e. the Indian government (OGD). These are just reported cases, and the real numbers are much, much greater than what we will be seeing (Shrivastava). To give you an idea, in Maharashtra alone, the number of reported cases in 2008 was 20,000 and in 2012 it was 42,000 (OGD).

“The only way people can understand data is to visualize it. That's how our brain works,” says Roman Stanek founder and CEO of GoodData, a company that provides data-visualization services (Forbes). Unlike numbers, colors immediately capture your attention and focus it where it's needed (Forbes). Information is powerful, but only if you understand it. Data visualization speaks to the human need to understand things visually (Forbes). Sonification is the use of non-speech audio to convey information or perceptualize data (Kramer). Pia Blumenthal says that because of our human ability to identify simultaneous changes in different auditory dimensions, we can integrate them into comprehensive mental images (I CARE IF YOU LISTEN). Sonification can be used to exploit this for the purposes of identifying trends and patterns in large sets of data (I CARE IF YOU LISTEN). And when paired with data visualization, sonification can provide a more holistic approach to exploring information (I CARE IF YOU LISTEN).

I want to make a website to bring awareness to the problem and let people see the real numbers behind what is happening. The following sections will include a history of data visualization, sound, the importance of each, my implementations, and my findings. I was fortunate enough to present this piece at the California Institute of the Arts' Digital Arts Expo 2016 (Digital Arts Expo 2016), and as a result I was able to talk to people about their views on the website, and further add to my implementation of the idea.

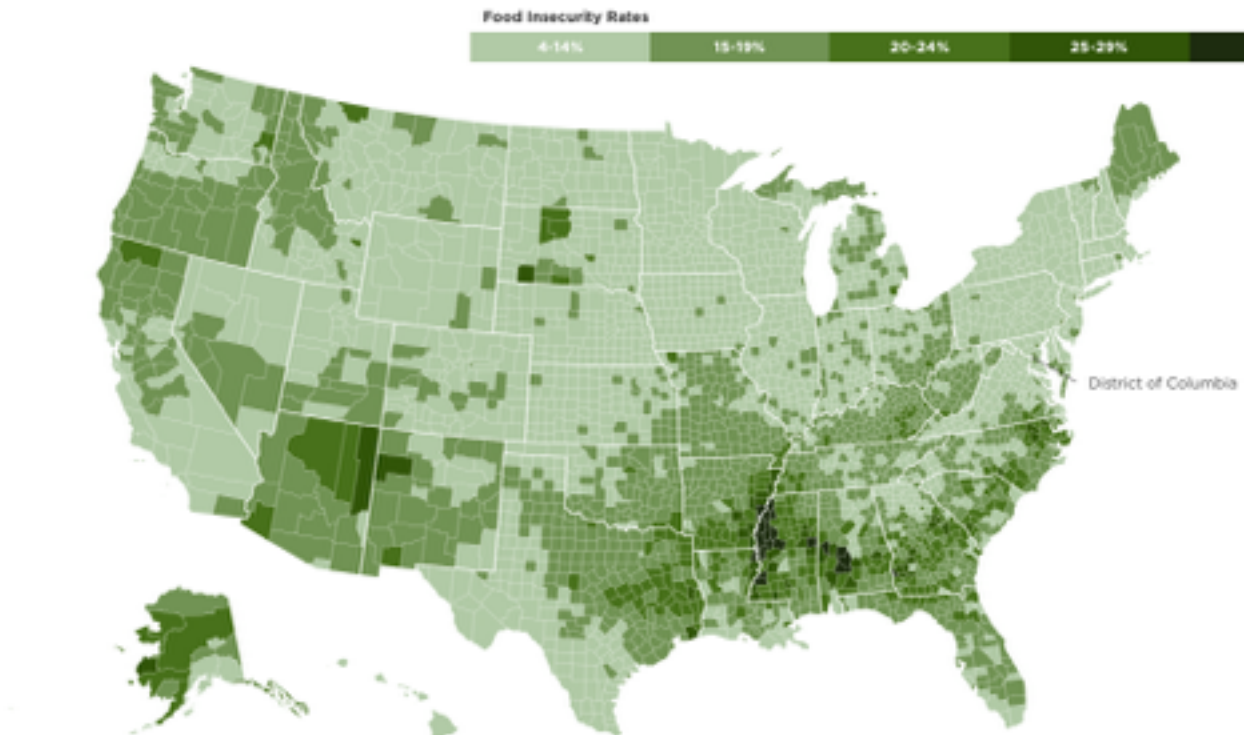
Why Data Visualization

Data visualization is viewed by many disciplines as a modern equivalent of visual communication (Friendly). It involves the creation and study of the visual representation of data, meaning "information that has been abstracted in some schematic form, including attributes or variables for the units of information" (Friendly). It enables decision makers to see analytics presented visually, so they can grasp difficult concepts or identify new patterns (Friendly). With interactive visualization, you can take the concept a step further by using technology to drill down into charts and graphs for more detail, interactively changing what data you see and how it's processed (Friendly).

Data visualization is both an art and a science (Friendly). The rate at which data is generated has increased, driven by an increasingly information-based economy (Friendly). Data created by internet activity and an expanding number of sensors in the environment, such as satellites and traffic cameras, are referred to as "Big Data" (Friendly). Processing, analyzing and communicating this data present a variety of ethical and analytical challenges for data visualization (Friendly). There are many examples of effective data visualization bringing about awareness. Here are two examples:

Feeding America has a great deal of data on food insecurity rates across the U.S. In 2011, they used this data to create an interactive "Map the Meal Gap" presentation that allows users to see and interact with county-level food insecurity data (Feeding America).

This is an excellent example of data visualization, and the organization has used it for advocacy and awareness purposes (Feeding America). When we compare their 2014 and 2015 financials, we see an increase of \$100 million in revenue (Feeding America). Of course, not all



of the increase can be related just to effective data visualization, but in 2014 there were fewer accessible data based projects for the public to see.

World Wildlife Fund, on the other hand, has created a series of infographics to help tell the story of what is happening in the Coral Triangle region of Southeast Asia (WWF-US). These graphic assets help raise awareness about the threats to a region that is not familiar to a large subset of the U.S. population (WWF-US). Their revenue also increased by \$200,000 between 2015 and 2014 (WWF-US). Also something worth noting is that they even changed the way they displayed these financials, and in 2015 they created graphs on their website to show the data.

History of Data Visualization

There is a history of data visualization: beginning in the 2nd century C.E. with data arrangement into columns and rows and evolving to the initial quantitative representations in the

17th century (Interaction Design Foundation). French philosopher and mathematician René Descartes laid the ground work for Scotsman William Playfair (Interaction Design Foundation). Descartes developed a two-dimensional coordinate system for displaying values, which in the late 18th century Playfair saw potential for graphical communication of quantitative data (Interaction Design Foundation). In the second half of the 20th century, Jacques Bertin used quantitative graphs to represent information "intuitively, clearly, accurately, and efficiently" (Interaction Design Foundation).

John Tukey and more notably Edward Tufte pushed the bounds of data visualization. Tukey with his new statistical approach: exploratory data analysis and Tufte with his book "The Visual Display of Quantitative Information", the path was paved for refining data visualization techniques for more than statisticians (Interaction Design Foundation). With the progression of technology came the progression of data visualization; starting with hand drawn visualizations and evolving into more technical applications – including interactive designs leading to software visualization (Interaction Design Foundation).

Programs like SAS, SOFA, R, Minitab, and more allow for data visualization in the field of statistics. Other data visualization applications, more focused and unique to individuals are programming languages such as D3, Python and JavaScript that help make the visualization of quantitative data a possibility. Because they use widely used modern languages, these tools are accessible and easy to work with. For example, D3.js has a large open source community where users can share example code, and graph creations online (Bostock).

Why Sound & Related Works

Sound is capable of producing powerful reactions in the listener - whether it's a sudden cold sweat caused by a snake's warning hiss, or the uncontrollable grin as a favorite song from our youth comes on the radio (Psychology of Sound). According to Julian Treasure, there are four major ways sound is affecting you all the time (Treasure). First is physiological. Loud, sudden sounds give us a shot of cortisol, which is our fight/flight hormone (Treasure). Conversely, the sound of surf has the frequency of roughly 12 cycles per minute (Treasure). Most people find that very soothing, and, interestingly, 12 cycles per minute is roughly the frequency of the breathing of a sleeping human (Treasure). There is a deep resonance with being at rest (Treasure).

The second way in which sound affects you is psychological (Treasure). Music is a powerful form of sound which effects our emotional state (Treasure). Natural sound can do that too (Treasure). Birdsong, for example, is a sound which many people find reassuring (Treasure). The third way in which sound affects you is cognitively (Treasure). For example, we find it difficult to hear people talking at the same time (Treasure). We have a very small amount of bandwidth for processing auditory input, which is why noise is extremely damaging for productivity (Treasure).

The fourth way in which sound affects us is behaviorally (Treasure). With all that other stuff going on, it would be amazing if our behavior didn't change (Treasure). So, ask yourself: when you're driving and listening to electronic music, are you ever going to drive at a steady 28 miles per hour? At the simplest, you move away from unpleasant sound and towards pleasant sounds. If you were to listen to a jackhammer for more than a few seconds, you'd feel uncomfortable; for more than a few minutes, you'd be leaving the room in droves (Treasure).

For people who can't get away from noise like that, it's extremely damaging for their health (Treasure).

And that's not the only thing that bad sound damages. Most retail sound is inappropriate and accidental, and even hostile, and it has a dramatic effect on sales (Treasure). They are losing up to 30 percent of their business with people leaving shops faster (Treasure). It is clear that there is a strong correlation between sound and the brain (Treasure).

According to Brian Foo, a programmer and an artist, unlike charts or visualizations, music is abstract and not that great at representing or communicating large sets of data accurately or efficiently (I CARE IF YOU LISTEN). However, music has a few clear advantages. First, it has the ability to evoke emotion and alter mood which can help the listener understand data intuitively and suggest how they should feel about the data rather than just communicate the data itself (I CARE IF YOU LISTEN). Also, music is beneficial to the creator since they can curate a temporal experience for the listener that can be consumed casually and viscerally (I CARE IF YOU LISTEN). In contrast, a visual chart can be navigated in many ways and usually does not impose a particular narrative structure (I CARE IF YOU LISTEN). Lastly, music gets stuck in the listener's head. If one can attach meaning or data to that music, perhaps those things (e.g. economic issues, environmental issues, personal stories, etc.) will get stuck in the listener's head as well (I CARE IF YOU LISTEN).

Brian Foo also created Data-Driven DJ, which is a series of music experiments that combine data, algorithms, and borrowed sounds (Data-Driven DJ). "My goal is to explore new experiences around data consumption beyond the written and visual forms by taking advantage of music's temporal nature and capacity to alter one's mood (Data-Driven DJ). Topics will range from social and cultural to health and environmental" (Data-Driven DJ). An example of one of these experiments is 'Two Trains: The Sonification of Income Inequality on the NYC Subway' (Data-Driven DJ).

“The goal of this song is to emulate a ride on the New York City Subway's 2 Train through three boroughs: Brooklyn, Manhattan, and the Bronx. At any given time, the quantity and dynamics of the song's instruments correspond to the median household income of that area. For example, as you pass through a wealthier area such as the Financial District, the instruments you hear in the song will increase in quantity, volume, and force. Stylistically, I want the song to exhibit the energy and orderly chaos of the NYC subway system itself” (Data-Driven DJ).

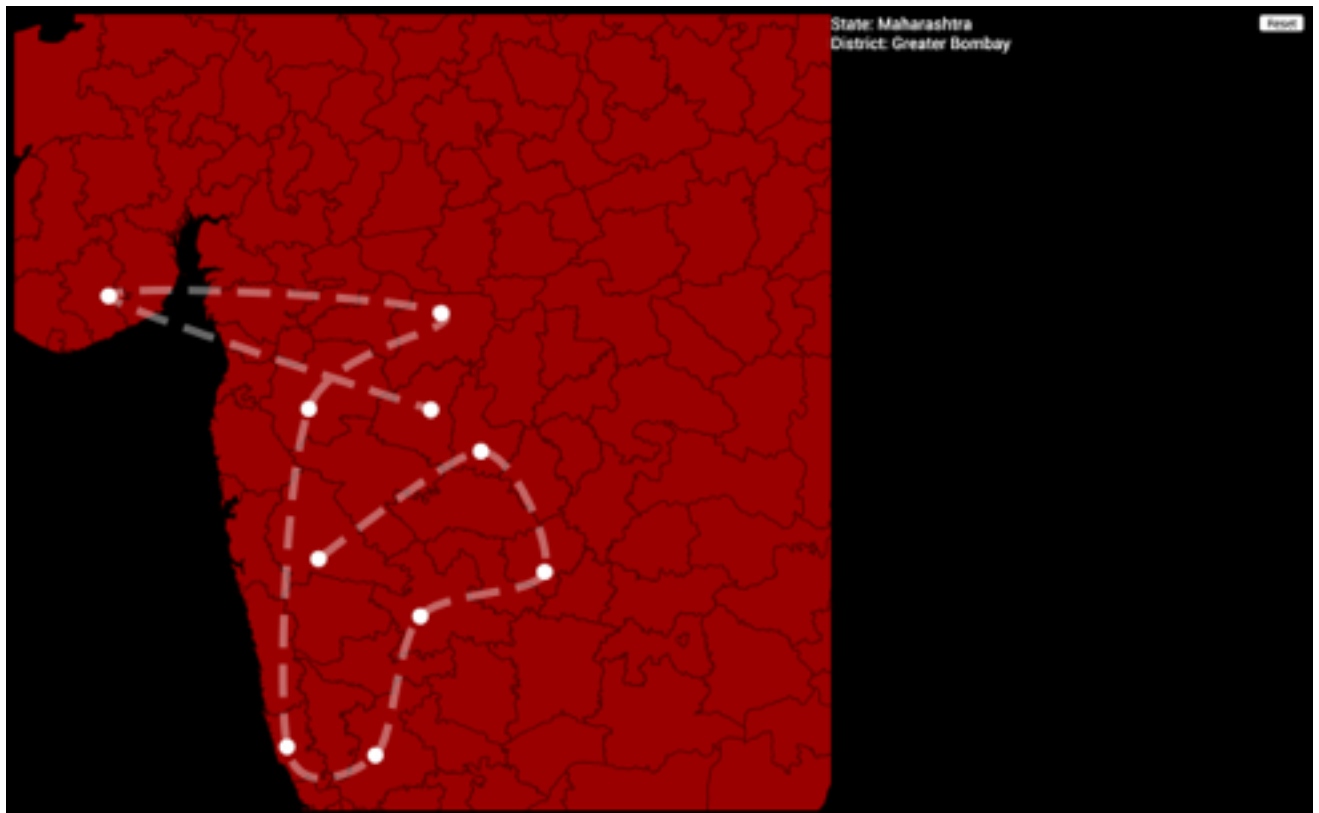
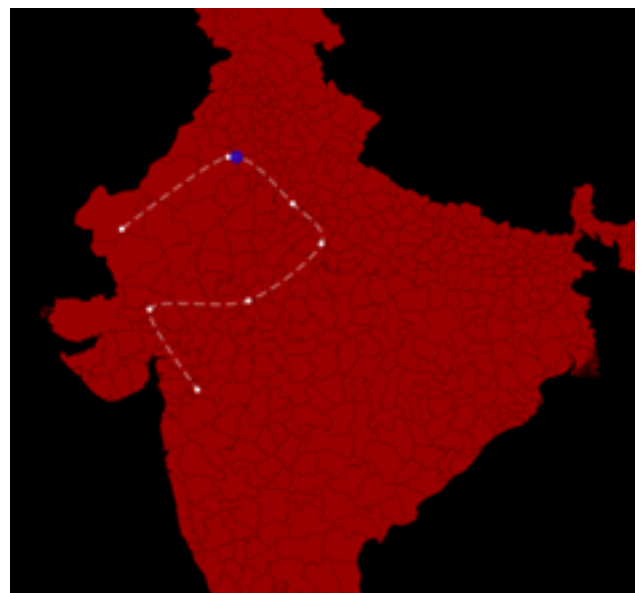
Implementation

To allow people to easily access the data, I created an interactive website. The data itself was stored in MongoDB, and I used Ruby on Rails to generate controllers for each dataset so I could fine tune what data I was querying at specific times. I used MongoDB, because of its speed with JSON objects, and querying abilities. The visuals were created in d3.js, an interactive javascript library used for data visualization. I used the p5.js sound library to abstract the WebAudio API, so I could focus on creating the soundscape to fit the visualization.

Visually, I went with the analogy of a tourist traveling through India via the railway system. I chose this because India has the largest railway system in the world, and an average of more than 23 million people use it every day. The user can create their own path along the districts, and the camera follows these paths while displaying relevant data to that district. The sound also changes based on the time it takes to travel. The sounds were recorded in India among crowds, which represents how unspoken these numbers are.



At first, I started off drawing just the map of states, but then progressed to drawing the districts as well. This gave me finer control of the data, as most of the documents I received from the government were indexed by district. Then I began creating SVG layers to represent the paths, and the train. I was able to allow the path to be drawn just on the map, by running the path drawing function within the mouse click listening function. I then implemented a reset button, to let user's draw new paths without having to reload the entire map. I also created a start button to draw the train once the user was done creating their path.



However, after implementing the camera translating (zoom) ability on the SVG layer, and after creating a camera that moved with the train, I saw significant performance reductions in the frame rate of the app. I still needed the map of the districts, to query the database, however it wasn't necessary for the end user to actually see every district line. In order to circumvent this problem, I decided to draw a transparent layer of the district map to help with querying, while the user would still see the map of the states of India. I also made the speed of the traveling train linear, and even scaled it based on how short or long the path was.

The next big issue was getting the relevant query to trigger whenever the train would pass over a station. I ended up using AJAX requests to POST to the database whenever I needed new relevant data. Once the data and visual loop was complete, I moved onto the sound aspect.

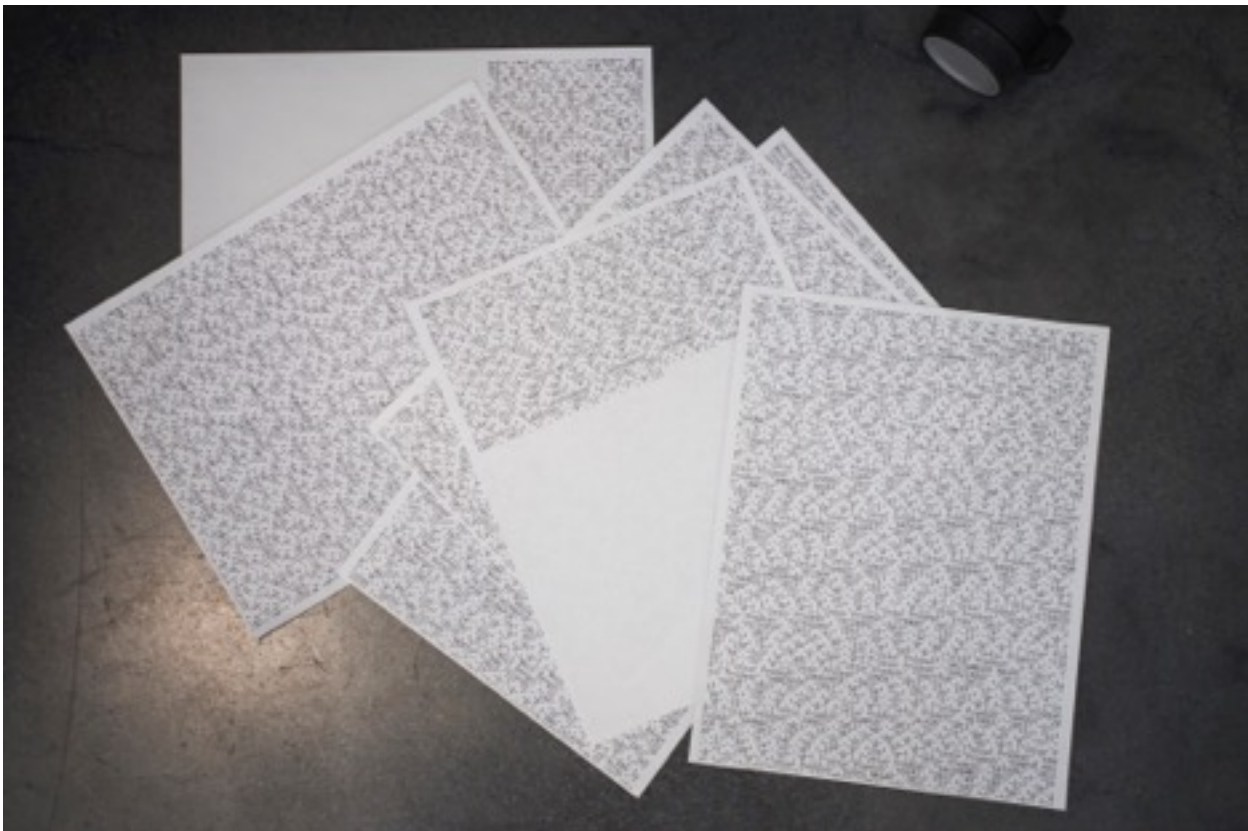
I was fortunate to visit Mumbai while I was working on the website, so I decided to take my portable recorder and get authentic sounds of crowds in India to use for my app. I wanted crowd and ambient noises, so I went to three different places to record. The first was Chhatrapati Shivaji Terminus, or Victoria Terminus (VT Station), as the locals call it. VT station is one of the busiest train stations in Mumbai, and I wanted to keep with the railway representation, so I had to record there. The second place was called, 'Bhaji Galli,' or 'Vegetable Street,' which is a bazaar style vegetable market. Finally, I recorded sounds of birds singing early in the morning, outside of my window at home.

After recording, I created a mix which had all three recordings layered on top of one another. I then created a lowpass filter in p5.js which would be active before the user started to use the app. When the user finished drawing their path, and as they would travel along it, the low pass filter would suddenly open up, giving more clarity to the sounds. Then as they reached the end of their path, the filter would close back down, muffling the sound and making it die back

into the background. I chose the filter because it signified that when users were discovering the data, there would be no obstructions to the sound, and everything would be clearer. I mixed the crowd sounds together, because I wanted to show that even when the sound was clear, it was still hard to hear what you wanted to hear, because of the number of sounds going on. This was intentional to represent that the data is inaccessible and even if you can get a hold of it, actually finding what you want is difficult and time consuming.

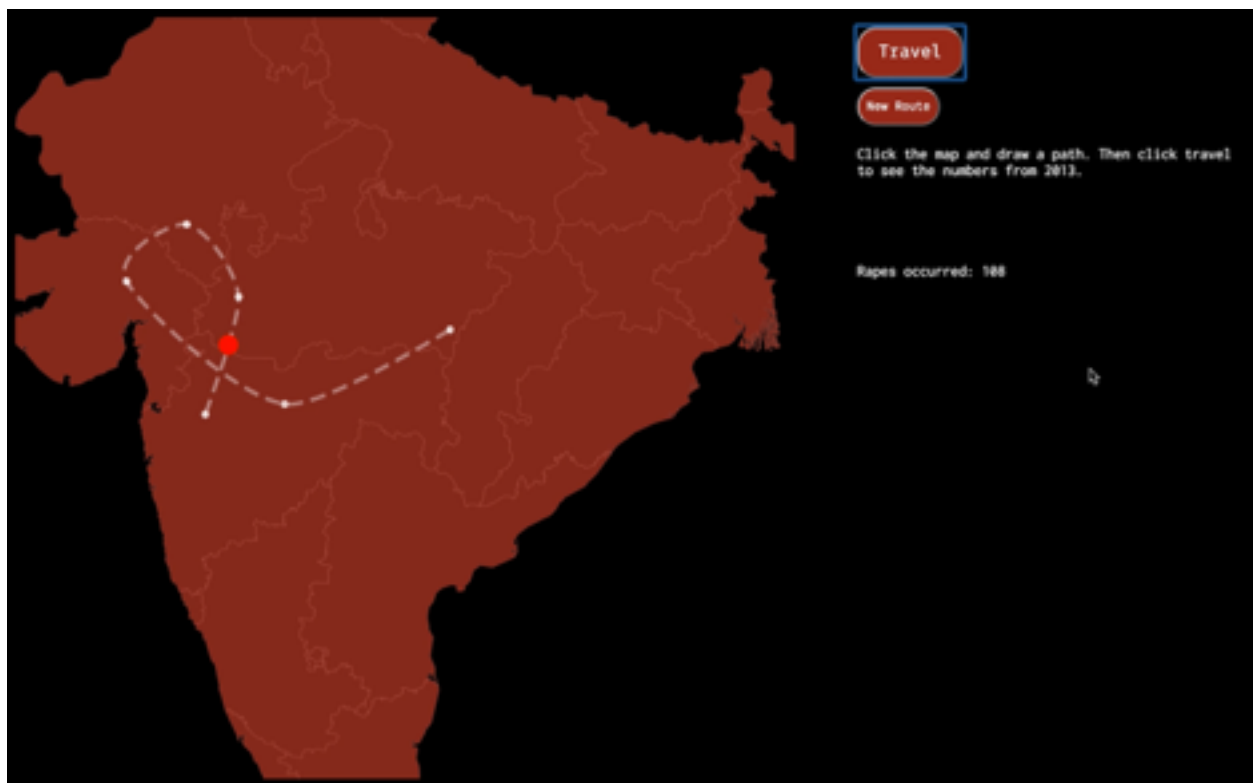
Results

The process of creating this app has not only increased my own technical abilities as a programmer, but it has also opened up my eyes to the reality behind what is happening in India. In order to show how inaccessible the format of the data was, I printed the raw JSON & CSV text from the files that the government would allow people to download:



Some of the sheets, I let fall to the floor, indicating that the data was going unseen, and few people actually take the time to find the information they are looking for.

I was able to speak to five different people about their thoughts on the website at the event. They were fascinated by the idea and the intent behind the project and wanted to learn more about it. They asked a wide range of question from things like, ‘What do these numbers really mean?’ to ‘How can we reduce the occurrence of rape in India this?’ to ‘Does the western media dwell on these numbers? Is India really that bad?’ They even acknowledged that the data looked unreadable, that these were only numbers till the year 2013, and that most cases aren’t even reported — all without any prompt from my side. All these questions were relevant to the outcome of the project, because my intent was to start a conversation between people and bring awareness to the issue.



Conclusion

In my mind, I had succeeded in my goals to get people talking and understand what is really going on in India's government. My idea was to have an artistic representation of the data in the form of an interactive website. Users were be able to draw paths through the map of India, almost as if they were drawing their own train stations across the country. The app would then slowly follow their path, and as they would pass over their created 'stations,' the relevant data for that district would be displayed. There was also an interactive sonified aspect, which immersed the user in the experience. The sound moved from the background to the foreground, as the user journeyed through India.

I have started a project that definitely needs to be more fleshed out and implemented. Not only do I need to add more data to the app, but I also need to make it more interactive in other more creative ways. For future work, I want to continue connecting more relevant data based on location. I also want a more robust implementation of the sound, and have the app start with pre determined routes. Watching people interact with the app, there is not much indication that can interact with it by clicking the map, and I would want to focus more on the user experience to allow people to discover the data easier. I also want to create a new page for people to see all the data just by looking at traditional graphs. This page would also ideally have statistics from other countries as well, so that the user can compare data between them. I believe I have started a conversation and kindled a sort of flame to bring awareness to this issue, now the only thing left to do is grow the fire.



Bibliography

"Open Government Data (OGD) Platform India." Open Government Data (OGD) Platform India. 7 Apr. 2015. Web. 09 May 2016.

Crime and Control in Comparative Perspectives: The Case of India. Shrivastava, R.S. (2011). De Gruyter. p. 190

"Big-Data Visualization Is Worth A Thousand Spreadsheets". Forbes. Forbes Magazine. Web. 09 May 2016.

"Auditory Display: Sonification, Audification, and Auditory Interfaces." Gregory Kramer. Santa Fe Institute Studies in the Sciences of Complexity. Proceedings Volume XVIII. Reading, MA: Addison-Wesley.

"Milestones in the history of thematic cartography, statistical graphics, and data visualization". Michael Friendly. 24 Aug. 2009. Web. 09 May 2016.

"Map the Meal Gap." Feeding America. Web. 09 May 2016.

"2015 Annual Report" Feeding America. Web. 09 May 2016.

"WWF-US Annual Report 2014" World Wide Life Foundation. Web. 09 May 2016.

"Data Visualization for Human Perception". The Interaction Design Foundation. 21 November. 2013. Web. 09 May 2016.

"Blocks". Mike Bostock. 4 May. 2016. Web. 09 May 2016.

"Impact of Sound on the Brain." Psychology of Sound. Web. 09 May 2016.

"The 4 ways sound affects us." Julian Treasure. 23 Oct 2009. Web. 09 May 2016.

"5 Questions about the Sound Visualization & Data Sonification Hackathon - I CARE IF YOU LISTEN." I CARE IF YOU LISTEN. 2015. Web. 09 May 2016.

"Data-Driven DJ." Data-Driven DJ. Web. 09 May 2016.

"Data-Driven DJ I Two Trains." Data-Driven DJ. Web. 09 May 2016.

"Projects." Digital Arts Expo 2016. Web. 09 May 2016.