Voice in the Data

MRP by Jacob Cram at York University

Contents

0 - Abstract

This project is an investigation into the use of data and data visualization in the construction of a documentary. This is done through the creation of a series of three data visualizations based on my own SMS history and photos, gathered over the course of 7 years. With each visualization building upon the emergent narrative, being careful to let the story that is in the data come out on its own, and not placing my own memories over it. The stories that are uncovered in the data are not a perfect replication of the past. There are pieces missing, and some that are out of place. There is an incredibly detailed depiction of what is in the data, but it's just that not everything is there. Using this format to construct a narrative lead to interesting results and insights. Pushing further it would be interesting to use data from more sources, like close friends, and comparing the two to see what story comes from the both combined.

1 - Introduction

Data has become one of the most valuable resources in our world. It powers almost everything around us, and is being created at a scale we cannot properly comprehend. We can't go through life without creating a digital footprint of some kind. But it's more than just a set of numbers, it catalogues everything in our lives. Our memories are contained to some degree within this data. While memories can change and degrade with time, our data does not. It holds an exact, if imperfect version of our

lives. What story is hiding within this data? And how can it be told through a documentary format?

There is a strong feeling of nostalgia that comes from going through an old photo album. Seeing these scenes frozen in time, and experiencing the emotions that went along with them all over again is a unique experience. With the advent of modern technology, this has become even more prominent. Almost every moment of our lives is recorded in data, through the photos we take, and the messages we send. Reading through old messages can feel at times like reading the words of a stranger, despite knowing that you wrote them yourself. Despite the data being your own, it feels as though it has the voice of someone else entirely. As though it had developed one of its own.

Drawing inspiration from the *Essay Film* [1] style, this project uses personal data in the form of SMS message history and photographs to experiment in the documentary medium. It explores the question; how can data visualization be used in the documentary format?

It is more than just a retelling of the past however. In a more traditional autobiographical work, there is an inherit bias that occurs from the impact that time and emotions can have on memories. Data is not effected by this, but itself has its own gaps and missing pieces. These gaps create a story that is different from the one a person remembers, and the one that is lost to history. It contains a story and a perspective that is it’s own.

2 - Theoretical Background

This project draws largely from three areas, the documentary practice of essay films, generative art, and data art. A secondary layer of inspiration comes from the phenomena of surveillance capitalism, and mass surveillance projects carried out by government organizations.

The beginning of this project came from building a storage server in order to always have access to my backlog of work, which at the time was scattered across a number of different hard drives. One of the first steps of this was putting all of my photos into one place, and having them be properly categorized. This lead to the creation of an experiment called *Gallery.CRM* [22]. It aimed to find a way to display every photo taken within a year, and categorize them through different visualization methods. Techniques explored in the creation of this project informed many of the decisions in both technical construction and aesthetics of *Voice in the Data.*

Luc Courchesne's *Naked in Paradise* [2] is a piece in a very similar vein. Displayed in 3D space, the piece exposes the entirety of Courchesne's works, allowing the user to float between them, and view each of them in reference to each other. It serves as an archive that is more than just a file structure, but a fully explorable space. This idea of creating an archive of data and projects is what sparked the idea for *Voice in the Data* initially before the thought of the story within that data came to be.

2.1 - The Essay Film

Following this there were a number of experiments with these images, but none that lead anywhere further. It was at this time that I began a course looking at the documentary style of the Essay Film. Though I had seen it before, the film *Sans Soleil* [3] stuck out to me in particular. The story being told is meant to be a little obscure, and unsure. There are glimpses as to what is happening within it, little slivers of messages, and of people. But because of a lack of connection to the source material, the viewer builds their own story walking through the exhibition, finding links, and seeing common names. The film uses the Kuleshov Effect [4] to give the illusion that everything going on in the film is connected, and all part of a larger story, even though that connection is totally manufactured.

Shortly following that was *Sea in the Blood* [5] by Richard Fungoes. The film details Richard's experience with thalassemia in his sister, and AIDS in his partner. The way the film presents itself with layered narratives happening at once is fascinating, with the narration, image, and overlayed text all telling versions of a similar story unfolding. Each speaking a truth, but all a slightly different version of that truth.

These films both handle ideas of varying truths in our memories, and our stories, and reminded me of another short film, *My Dead Dad's Porno Tapes* [6] by Charlie Tyrell. The film tells the hidden story of his late fathers life through the objects he left behind. Showing the story that Charlie thought to be the truth of his father, and how different it was to the one that he only uncovered after he was gone.

2.2 - Data

The combination of looking through all these old photos and the ideas of memory and truth that these films contain put me on a train of thought of what stories and truths could be found through these photos. The work of Lev Manovich was interesting through this phase. He writes about how data and computation can be used to analyze large social and cultural phenomena in his piece *Representing Phenomena as Data* [8]. He presents a series of decisions that help shape how this data could be analyzed, and presented.

Discussing what should be considered the boundaries of the phenomenon, what metrics should be used to represent it, and exactly what properties of each data point should be included.

In the text he leads through a series of questions surrounding how the data should be worked, defining the boundaries, representation of that data, and which properties are relevant to the phenomena. This discussion of properties became particularly relevant, as Voice in the Data sheds the actual data, and focuses on the metadata of the photos and messages. This changed the nature of the project from being an exact retelling to being more abstract, and representational. These questions helped to guide me along the path to this project. Leaning away from just purely being a series of data visualizations, and into creating a narrative in the same way the Essay Films do.

Using my personal data for a project isn't something that is new to me. *Text Based Life* [9] is a project from 2017 that was based on my personal text message

history, using the frequency of keywords used in conversations with my 10 most text contacts to create a set of 4 visualizations. While an interesting project to look at myself, the most interesting part was showing it to people who didn't know anyone on that list, and hearing them theorize about who these people were to me. With such a limited dataset it was interesting to see what people could gather about not just my life, but the lives of these other people as well.

Text Based Life [9] accidentally demonstrated very well how big of an impact, and how valuable metadata can be. Metadata is a form of data about other data. So for example while a text you send may be data, the information about when it was sent, to who, and how long the message is is all metadata. [10] The distinction between them becomes very important when looking at projects such as ones uncovered by Edward Snowden in 2013 [11]. It may seem at first that the collection of metadata is far less of a concern than the actual data, but it can give even more insight into your life and relationships than the data itself could. Especially when given a significant amount of time for collection. Patterns can come to light that would have gone unnoticed. This is the reason the project focuses in on just 2 sets of data, messages and photos. While far more could have been used from different sources, I chose to limit it to those two to see what kind of story could be told just from a limited set of data. Just through the information around my messages, you could even approximate my location when I sent the messages to a fairly accurate degree.

Motivation for this project also came from a basis of 'taking control' of my own data. A large amount of it was gathered from sources like Google, who use that data to

build a profile of me to push personalized advertisements. David Carroll is a professor of Media Design at Parsons School of Design, and has been a figurehead for exposing the hidden systems that surround us, and use our own information against us in various ways. He discussing this emergence of Surveillance Capitalism in a medium article [12] describing the methods to avoid its effects, and was heavily involved in exposing the Cambridge Analytica scandal, where large amounts of user data was scraped from the site without the user's knowledge (and understanding) for the purposes of targeted political advertisements. [13] The data used for these advertisements only came from what was available on a persons Facebook page, and contained a shocking level of detail and granularity about a persons political leanings, their mental stability, and even down to their grocery habits. If my data is going to be used against me, and potentially others, I wanted to use it for my own gain as well.

2.3 - Bias in Data

Data and computation are often thought to be an unbiased and totally truthful version of information. Free from the prejudices humans tend to place on everything. This is a myth. Data can be just as biased as anything else. Facial recognition and image tagging have repeatedly had issues in this regard. [14] Because while it may be a computer generating the data, the collection of it had to be designed by a person. Algorithms can be just as racist and blind as the people who create them.

This was an interesting topic during the creation of *Voice in the Data*. However in the end the bias that was present in the data was past the point of being something that could be removed. Removing bias in this case is not an effort that would help the

project. In fact because it is about revealing the biases in the digital history we create, and in our own memories of the past, it was crucial to ensure that whatever biases were in the data remain there.

2.4 - UNMASK + Brutalism

My previous work, *UNMASK* [15], was specifically a study into how the practices and theories of Brutalist architecture could be adapted into the realm of digital media. During that time I developed somewhat of a framework for working in a “Brutalist” manner.

Brutalist architecture is an often misunderstood style. It’s most associated with large, looming buildings that make the viewer feel small and insignificant. Buildings that are concrete monstrosities, trying to be as unappealing as possible. There is some truth to these ideas, however the reality of brutalism is very different. The name itself is flawed in English. It seems to be based off the word ‘brutal’ when in reality it’s based on the French ‘Beton-Brut’ or ‘raw concrete’ [16].

Brutalism was born out of the modernist architecture, but went a step further in rejection of previous techniques. The goal of brutalism was to create buildings that were as honest as possible, and stripped of any kind of ornament [17]. Not function-over-form, but function-as-form. Key elements of the buildings function were used as centrepieces in the design, elevator shafts and stair wells became prominent in residential buildings, and ventilation was heavily featured on the exteriors. When

rebuilding cities following the Second World War, brutalism was used extensively as an ‘architecture of the people’ that was cost effective, and simple.

This stripping of any ornament is very similar to another child of modernism, minimalism [18]. At the surface their goals are the same, but they come at it in different ways. The best way to think of it is to picture a brick wall. In a minimalism mindset, you would paint the wall to be a consistent, flat tone, such as a consistent white. In brutalism, the wall would be left alone, as anything else would be a facade. Minimalism is simplification at all costs, brutalism is honesty at all costs.

Adapting this honesty into other fields leads to interesting results, but again, sometimes misguided. The first step is to identify the ‘bare materials’ of whatever medium you are producing for. The goal is to make these bare materials as unobscured as possible. Identifying them can be a process unto itself. For buildings it is the concrete they are made of, but for a website the first thought would be the HTML structuring the site, however this isn’t the true ‘material’ of what is being made. The material is the actual content of the website. Meaning that in a brutalist web design, the goal is to make the content of the site as unobstructed and readily available as possible.

The next stage of the process is a constant justification of everything you add to the project. Any decision needs a reason to be added, and a reason for being a part of what you are making. If there is no reason, it is just ornament, and should be stripped. This stage is often the hardest, as it forces you to constantly question what you are making, and potentially remove pieces that you are heavily attached to. But it is a

crucial piece of the process. It will lead to a much more clean and cohesive project at the end.

It is also very much worth noting that while I am a strong advocate for brutalism, and the radical honesty, I do not recommend this method of working for everyone, or for every project. It can take a heavy mental toll during creation. This can lead to breakthroughs, but it can just as easily lead to the joy of a project being torn out of it. Sometimes we create things for the ornament.

For *Voice in the Data*, elements of this brutalist process were used, but it was more sparing than in *UNMASK* [15]. Each of the visualizations used this brutalist process to justify their contents. This helped to let the data tell the story that was within it, rather than imprinting my interpretation on it too heavily.

This honesty lead me down a more personal path in the creation of the installation as well. I chose to take that as a philosophy and based the installation on my own struggles with memory distortion and destruction as a result of anxiety and depression. *Voice in the Data* exists almost as a response to that installation. Where it was about the destruction, and expressing the pain and anguish that comes with that, *Voice in the Data* is an effort to rebuild. Using the data that I have, I can put together the pieces that are left. But it is not an attempt to perfectly recreate the past. As said before, it is an attempt at seeing what exactly the story is within the data. If there are irregularities or errors within that data, those should not be corrected, or altered. The mistakes need to be included, otherwise it isn't the story that the data holds.

2.5 - Aesthetic

I often find myself drawn to more Brutalist aesthetics as well, and these would inform a number of points in the project, specifically in the creation of Contacts. The book *Form + Code* [19] was a source of a lot of early inspiration for the aesthetics of the pieces. The book covers the history and theory of using computers for the purposes of generative visuals, and the ways to make the computer do the heavy lifting. Image averaging is one in specific that was experimented with, but did not find itself in the final product. For the first of the pieces, Contacts, it drew me towards a more 'direct' style. Combined with the aforementioned brutalism, the intent was to show the information in the most clean and plain way possible.

Inspirations for the aesthetics of the piece didn't stop in the art world either. The concept of Punctuated Equilibriums [20] came later in the process, but gave an interesting light on how to progress the aesthetics between the pieces. Where Darwin's initial theories of evolution propose that evolution is a continuous process, with a species accumulating small variations over long periods of time until a new species is born, Nils Edredge theorized that new species come from the fringes of a population, rather than the main stream of it. With factors such as geography creating a new 'pocket' of the species where its evolution differs from the rest of the species. This idea took root mainly in the distinction and creation of Contacts and Galaxies. The two share a similar DNA, pulling from the same set of data, and same basic technological construction, but their evolution took completely different directions, leading to two heavily connected, but entirely different pieces. This theory was considered as a component for the timing of each of the elements, however did not make it to the final

iteration, choosing a more fluid and slow growth over time that seemed to suit the pieces better.

Vortex takes the majority of its aesthetic direction from a piece created in mid 2020. While it began as just an experiment in pushing the limits of TouchDesigner [22], *Gallery.CRM* came to be incredibly valuable in defining how to sort massive numbers of images in 3D space. It was a project with a similar goal of cataloguing photos, though it had a focus on examining a single year at a time. The images were all laid out in 3D space, and the arrangement of them could be changed based on the time the photos were taken and their colour data. It was the start of my experimentation with a virtual gallery space, and was the start of this style of piece with a blend between catalogue and artwork.

This pushed a more ‘naturalistic’ aesthetic for Vortex and Galaxies. Drawing less from the ideas of computation and more from the slow growth and evolution of nature through time. Those two follow a similar evolutionary path, but through their aesthetics rather than the data. Vortex came from the idea of having every image put into a slowly churning storm, and Galaxies was born of the same idea, but instanced for every contact. Each piece is a reflection of another in some form. Drawing inspiration from the Sonata [23], a style in music in which a long form piece is created from smaller segments repeated and reflected on each other.

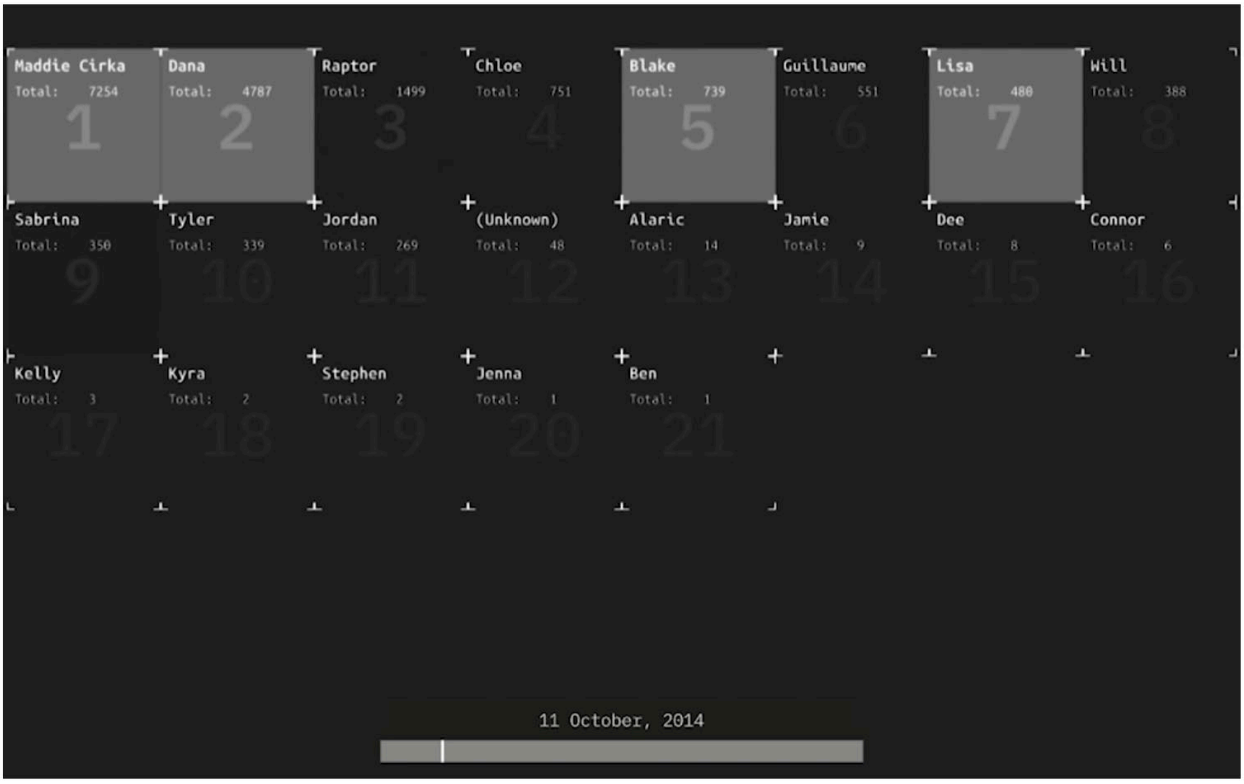
3 - Methodology

The process of creating this series of works was done primarily in segments, with a full focus being put onto each piece individually. Each was developed almost in a vacuum from each other. Only when each piece was largely complete were they combined a unified for the final installation appearance.

The final important part about working on each piece was to ensure that the data was being represented in the most clear way possible. While the goal is to create a documentary of my life through the use and visualization of this data, the aim is to tell the story that is within the data, and imparting my own biases and memories over that would change that. The story has to only be coming from the data, including what could be considered ‘errors’.

3.1 - Ethics

This project is by its very nature a fairly invasive one. Using personal data to generate visualizations has the potential to release very sensitive information. While the data is my own, it contains private conversations with people who may not want those conversations being in the public eye. In addition, the number of people and the time scale of the data makes it unfeasible, and in some cases impossible, to obtain permission to display these private messages.



In order to pursue this project in an ethical manner, none of the contents of any messages will be displayed for the pieces, and the last names of the contacts have been removed.

It feels very necessary to go further than is typically needed to ensure the privacy of all those involved, especially if this project is supposed to be somewhat of a show of how much information personal data can really hold.

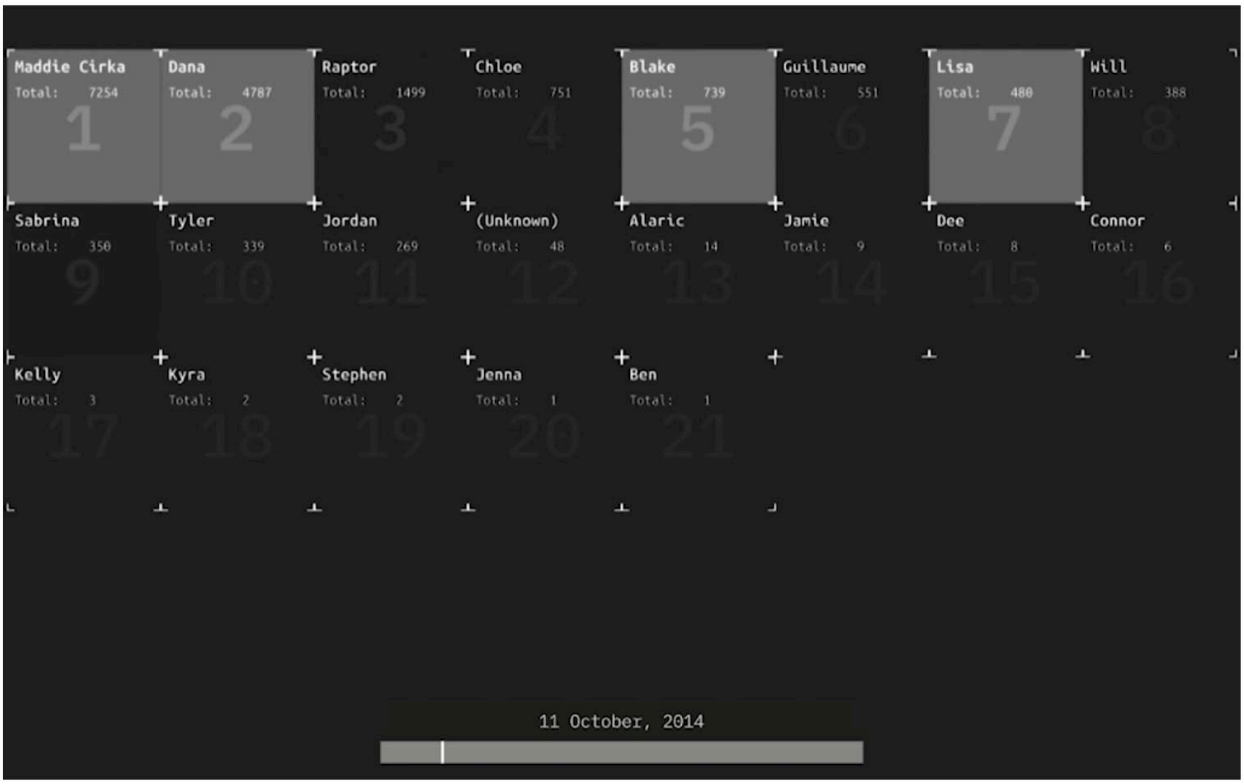


Figure 1 – Contacts, the first of the visualizations.

4 - Contacts

4.1 - Ideation

The first piece acts as the introduction to the installation. It is the most basic and straight forward of the visualizations, featuring a simple grid of every contact, in order of most to least messaged. As time moves forward, the contacts will show the number of texts sent and received, and will move around on the grid accordingly. Contacts will rise to the top of the list before descending back down into obscurity, and occasionally rising back to the top again.

It serves as a sort of exposition of what is going on. It explains who these characters you will be seeing are, and how important they are in the overall story.

4.2 - Technical Construction

The first stage of this development was parsing the data to all be in the same format, and be readable to use across the different pieces. The total for all of the SMS data being used was well over 100,000 messages, all with their own accompanying metadata. The sheer size of this data set required the set to be scaled down for development and testing, with the year 2019 being chosen for the purposes of development and testing.

The next stage was primarily formatting the data so it was easy to work with. While this is nearly half of the code created for this particular piece, it is the foundation of the entirety of the installation. Each of the pieces involve a large amount of code that is dedicated towards formatting the data in a manner that is optimal for that piece.

With all of the messages properly formatted, it was a simple matter of counting the messages sent and received by each contact, and formatting those into a table. This table is then output through a network connection to all of the other pieces, and onwards towards the visual component of this piece.

4.3 - Master Clock

In music production, a Master Clock [24] is used to keep everything in sync, and on the correct tempo. In order to keep all of the pieces in this installation working together and at the right pace, a similar concept was needed.

Unix time was chosen as the basis for this clock. It is a highly granular measurement, reading down the the milliseconds that a message was sent in the SMS data, and has very easy and universal conversions. The first iteration of this mechanic was re-purposed from a precursor experiment. However the requirements of this experiment were different from this project, and because of this it was experiencing issues with skipping numbers. The entire mechanic was redone in a much more robust fashion, however this was at the cost of performance, reducing the frame rate that the piece was able to run at, but not to an unacceptable level. This operates as the heart of all of the pieces, keeping them in sync with each other and running at a consistent pace.

Figure 2 – Vortex, the second visualization, based on photos.

5 - Vortex

5.1 - Ideation

The origins of this piece came from another experiment titled *Gallery.CRM* [22]. It was an effort to analyze a collection of personal photos and experiment with different forms of presentation for them, arranging each photo based on the average colour of the image.

The vortex motif was inspired to a degree by the never ending storm in Jupiter's 'Great Red Spot' [25]. A force of nature that is constantly churning in a timescale beyond our comprehension. This is what came to mind when I was attempting to

conceptualize these photos and memories in my mind. A level of sensory overload to a certain degree, with images being gathered and floating around so quickly it can be hard to understand what was just seen before another is occupying your view. It's a very powerful and chaotic force, but is also peaceful in a way. The patterns and movement as a result of the chaos being almost hypnotic.

The goal of this piece wasn't to make a point of the contents of the images, but more the presence and absence of them, with the content falling secondary to that.

5.2 - Technical Construction

While being the most simple of the visualizations in terms of technical complexity, it was the most challenging of the pieces to optimize. The initial version of the piece were attempting to display the full image in the vortex as it spins, allowing a viewer to actually see the image as it spins by each time. However in just the 2019 testing data set, the total amount of images was over 1000. While loading the first few images was relatively easy, as the visualization progressed, it became slower and slower.

This issue was compounded by how the timing on the master clock interacts with the piece. Handling the SMS data, each message includes a timestamp going down to the miliseconds that the message was sent at. This means not only that each image has its own unique identifier, but also that only one message will be loading per cycle. The photos do not have such granularity, only going down to the seconds that the images were captured. This meant that images that were captured in quick succession would

attempt to load all at once, causing a huge stutter as the program attempts to load the images with several bandwidth bottlenecks.

This was compounded by the images coming from multiple different cameras. In the 2019 data set, there were 3 cameras in total, but when expanded to the full range of years, that number grows to 7 cameras. Each camera has not only a different resolution, but can have a different aspect ratio. Meaning that the object each image was attached to in the 3D environment would have to have conditional sizing depending on their aspect ratio.

Unlike the previous piece which had most of the performance issues solved through splitting processes onto multiple CPU threads, there was little that could be split off into a separate process for this piece.

5.3 - Building the Storm

The placement of each of the images around the vortex is incredibly important to the piece. Their placement around the circumference of the vortex and vertically along it would be decided by the time that they were taken, but choosing the scale was important. The images had large gaps between when they were taken in spots, with over a month between captures in some places. While these ‘holes’ in the vortex were just as interesting in some ways as the spaces filled by images, their size needed to be managed. If the images are packed in too tightly, you won’t see them each appearing in, and if they are too far you lose the larger sense of time and scale that the whole vortex is meant to produce.

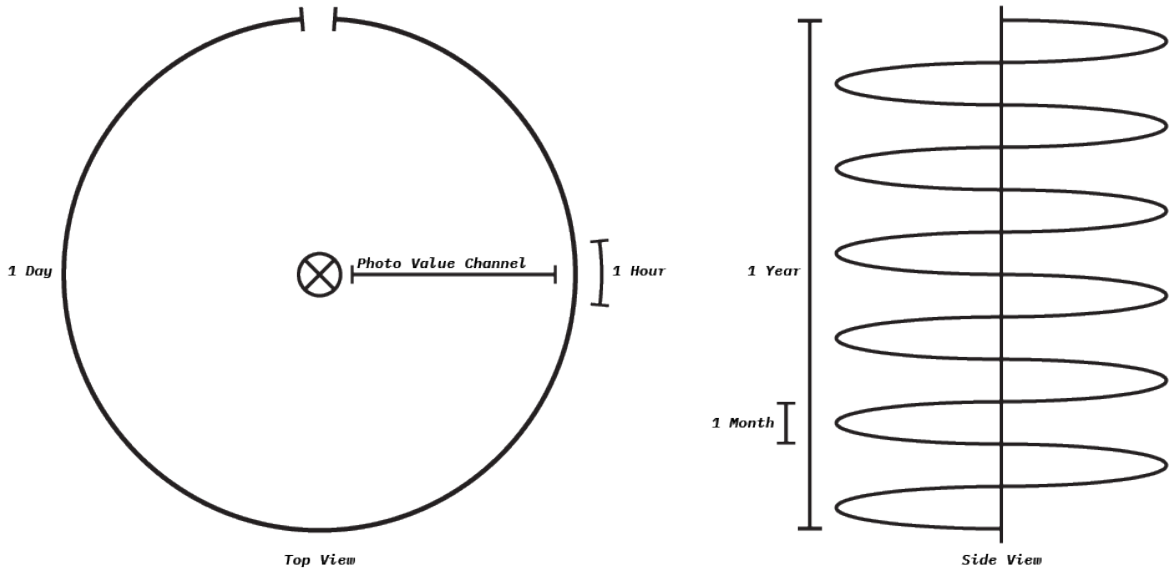


Figure 3 – A diagram of the construction of Vortex, depicting how the photos were placed in the storm.

It was chosen to have each ‘ring’ around the vortex represent 1 day, while the month and year in which the image was taken would decide its vertical placement along the vortex. This provided a good balance in focus on the most recently added images, and giving a context for the vortex as a whole.

Their position along the radius of the vortex was another key component. This initially was decided based on the overall brightness of the image, with darker images floating towards the edges and lighter being pulled towards the center. However the saturation differences in each of the images caused this layout to look far too noisy in the visuals. A combination of the saturation and the value was decided, combining the two methods to create a gradient from the center out towards the edges.

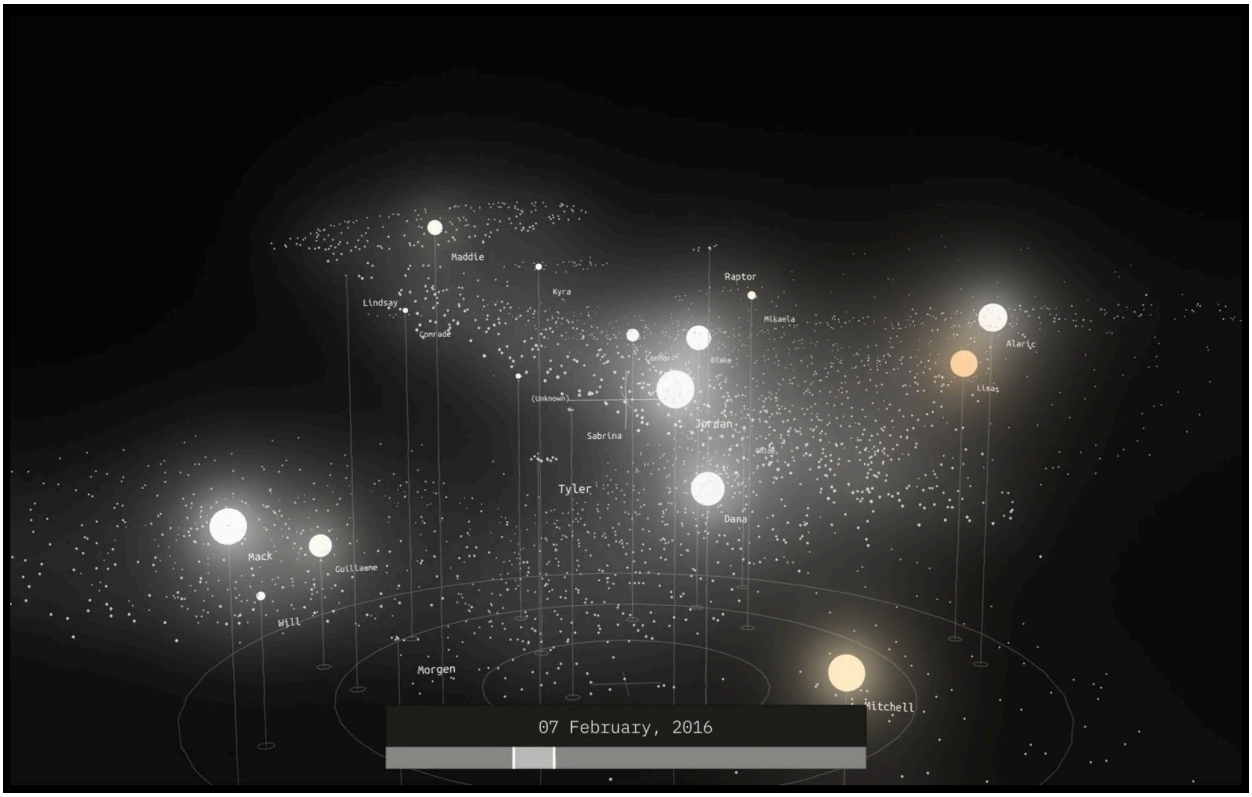


Figure 4 – Galaxies, the final visualization, based on slices of SMS history.

6 - Galaxies

6.1 - Ideation

The ideation of the final visualization took far more time than any of the others, and went through a number of iterations before arriving at the its final state. While the majority of inspiration and reference for the installation came from film, the progression to this final visualization came from music. Specifically the idea of the Sonata [23], which was previously discussed. Progression is created by reusing the same initial

ideas and patterns in different ways each time. Each movement becoming somewhat of an abstraction of the last.

While Contacts introduced all of the characters, and gave an overview throughout the entire span of time that the installation covers, it only offers the perspective of who was most contacted of all time. Leaving out the ebbs and flows of relationships, and people who come and go from life, coming in quickly and disappearing just as fast.

This piece is attempting to show that. It provides a more ‘in the moment’ view of the messages and relationships, shifting and adjusting as they evolve over time.

Due to privacy concerns, the actual body of the messages sent and received were scrubbed from the project, however the data could still be useful if obscured in the right way. Such as using an emotional analysis algorithm to determine the tones and emotions present in the messages.

6.2 - Swirls

Though the original idea was to have each contact be a line with particles appearing along that line according to when a message was sent or received, this was changed after creating Vortex.

Galaxies responds both to Vortex in its visual language, and Contacts in its goal in storytelling. It takes each and evolves them slightly, merging them together to create something altogether new.

Each message would be represented as a star within the galaxy of that contact. Swirling around in the center for a time before burning out, and fading away. Each message being given a colour according to their emotional values, making each galaxy have a unique appearance depending on the overall nature of the relationship. With stars of different colours replacing each other, showing changes as time goes on. Never looking quite the same moment to moment.

6.3 - Technical Construction

Despite the technically much higher impact of attempting to calculate the positions of over 100,000 particles that represent the messages, the more restricted time frame that this visualization covered made optimization fairly simple. The visualization is only covering 6 months worth of messages at a time, reducing the number of particles being rendered significantly.

The key component of this visualization was the movement of each galaxy. Nothing in space is ever stationary, everything is always moving and flowing, and the galaxies in this had to be the same. By generating random points within the volume of a sphere, each galaxy given a unique location within the universe. Each galaxy is then able to rotate around this point, with different rates for each of them, based upon a seed generated from the contacts name. Each one rotates slowly on all of its axises, with the Y (vertical) rotation moving the stars around with it.

The scale of each of the galaxies was then tied to the number of particles that were currently existing within it, shrinking into nothing when there were no particles. An

upper limit was added to prevent a single galaxy from every getting so large that it would dominate the entire frame and take over the visualization. This also prevented any anomalous numbers from causing too many issues.

6.4 - Emotion Analysis vs Sentiment Analysis

The emotional analysis of the messages was performed using the ParrallelDots Emotion Analysis API [26]. The API gives each message a ranking from 0-1 in each of 6 catagories; Happy, Angry, Excited, Sad, Fear, and Bored.

After the response from the API, each message is given an index based on which of the emotions had the highest score. Within each galaxy, there is a series of particle systems for each emotion, with the sum of all of them being used to calculate the size of the galaxy.

Each of these different emotions is assigned a different colour. This choice was based upon the colours that exist in stars in our universe, being limited to a natural spectrum that occurs based on the size and temperature of a star. [27]

Worth noting is the difference between the more commonly known Sentiment Analysis [28] versus Emotion Analysis. While they appear to be very similar, there is a higher level of granularity in emotion analysis. Sentiment analysis aims to find the overall positive to negative sentiment of a piece of text. It is useful for determining the perception of a product by scanning social media, and gives a simple single number reading of how positive or negative a sample of text is. Emotion analysis dives deeper, looking into the specific emotions in the sample of text, and breaking them out into

different amounts. Sentiment is generally more accurate, as it is looking for a much more loose structure, with less variables, but does not provide the detail needed for this.

6.6 - Visual Effects

The previous visualizations did not require an additional layer of post processing effects, as they would only obscure the information trying to be conveyed. This piece however required a variety of blurs and trail effects to fully complete the appearance of the galaxies. The blur makes each individual star less pronounced, and gives it the appearance that they are each radiating light outwards. As with real galaxies, this gives a more uniform appearance to the system, making it appear to be more unified than just particles orbiting an object.

6.7 - Revisions

While the majority of the revisions to each of the pieces came following the initial test installations, Galaxies went through a number before this. To the point where there is another piece entirely that was slowly scrapped as it was updated. The original display of it felt very ‘still’. Despite the amount of movement with the particles and galaxies moving, it still felt like everything was very stuck in place. By adding an orbit to the camera, it brought the level of motion to the level that was imagined. The camera always faces the center of the ‘universe’ while it rotates slowly around it, with a slight rise and fall as well to give a different view every cycle.

6.7.1 - Pulses

It was also decided to add a ‘pulse’ from the centre of the galaxies every time a new message was sent or received. This helped connect it a bit more with Contacts, and made the installation feel more cohesive. The pulse was originally much more subtle, with just a soft push each time a message came in. This was changed to a much more noticeable and attention grabbing pulse in later versions.

6.7.2 - Emotions

Though the original intention behind Galaxies was to have each ‘star’ be colour coded depending on the analysis of the API, this was scrapped due to the number of messages that would need to be processed, and the budget alloted for this. Though it was tested with the limited dataset from 2019, and each of the messages from that year were analyzed, the number of messages from the full 7 years meant that for this to remain in budget, it would take 3 months to process each of the messages.

The colours for the ‘stars’ were maintained, and for the 2019 year are coordinated to the emotion analysis, but for the other years are randomized.

6.7.3 - Global Clock

To further the connection between the pieces, a global clock was added to the bottom of each of the pieces, showing the current time along with a progress bar of how far through the data the time currently was. This gave more context to what was happening in each of the pieces, as well as better explaining the scaled time that is occurring in the pieces.

Galaxies was given a clock differing from the others, showing the range of time that is represented by each of the galaxies.

6.7.4 - ‘Baking In’ Functions

With a project like this, there are a lot of moving pieces, all of which add up, and can lead to major performance issues.

While each message of the piece was represented by a row of a table, it is no as simple as just scrolling through this list as time goes on. This would mean each message comes in with even spacing between them, which is not how the messages were originally sent. Each message has a different amount of time between it, and the one before and after. To represent this, a component was needed to count through the Unix timestamps of each message, checking each update to see which messages had a timestamp value less than the current time.

The issue was that this had to be calculated for each message every frame. Leading to major performance issues as the piece went on. Vortex avoided this by offloading content that was no longer visible on screen, and Galaxies by the nature of the piece only had 6 months worth of messages loaded at a time. However Contacts required the full history of messages.

Multiple attempts were made to make the process more efficient and keep it in real time. However each attempt proved to be less efficient than the last. What was chosen to be done in the end to keep the visuals of the piece functioning was ‘baking in’ [29] the progression of this time by rendering the line index of the most recent message

into an audio file. Touchdesigner would read this file, and play it back in the correct time.

6.7.5 - Controls

Due to current circumstances in the world, the installation will not be viewed in person, however a method of controlling the installation adds a great deal to the experience, and the connection that a view has with it. In past projects, I have been hesitant to add too much viewer interaction. However during development there was a slider to exactly control the passage of time that was engaging to move.

With the motif of circles through the installation, it was decided to keep with this and use a rotary encoder so the viewer could continually turn the dial to move through time both forwards and backwards, with the installation moving forward as normal when the dial is stationary. A ring of LEDs provide feedback for where through the time line the installation is currently at.

6.7.6 - Digitizing for Current Times

This project was created entirely within the global COVID-19 pandemic, and because of this was worked on with very few eyes seeing it during development. This alone was a very different process, however it also means that an exhibition of the installation can’t happen as it normally would. Several ideas for a virtual version were examined, and in the end it was chosen to create a web version with the A-Frame [30] tool from Mozilla. While the online version is not a live presentation, and does not allow

for user input or control, it does illustrate what the installation is about, and provides a form of it that can be viewed by anyone, no matter where in the world they may be.

7 - Results

Each of the pieces was displayed on a large format projection screen around 3 walls of a room, with the viewer able to walk around and view each. The flow of the ‘story’ follows from left to right, starting with Contacts providing the basis and giving an overview of each of the people featured in the installation, moving into Galaxies, which provides more context towards the relationships, and how they change, appear, and disappear over time, and ending with Vortex, providing a conclusion in the visual fragments of memories floating through space, providing glimpses into the reality behind the messages.

Located in the center of the room is the control dial, which a user can move to view specific times in the installation, rewinding and fast-forwarding through time as they wish.

The scale of the pieces provided a stark change in how they were perceived. During the majority of development they were on small screens where the focus was all on a single piece at a time. The large scale gives a sort of sensory overload even with the relatively simple visuals of Contacts.

7.2 - Self Epiphanies

Creating a project about your own life is a very self reflective exercise. While during the creation of each of the pieces I wasn’t directly seeing the data all the time, it

was required to have the full set visible in order to debug issues, and make sure everything was running the way it was intended.

The intensity of working on a project of this scale puts your mind into a kind of flow state. You turn off certain pieces of your mind and are entirely focused on the goal in front of you. Those familiar with the game Tetris have likely heard about this state. [31] Reaching this flow for this project became difficult at many points though, as while I was working I would catch a glimpse of a message, and scroll a little to find the context. I would end up reading whole conversations I had totally forgotten about, with people I forgot I used to be so close to.

The COVID-19 pandemic was also taking place during the entirety of the creation of this project, and the majority of that time was spent in Toronto during one form of lockdown or another. Spending a great deal of time alone with these messages allows for a lot of time for self reflection. While this can be argued to be healthy, at some point it becomes almost regressive. I found myself dwelling on the past far too much, and needing to remind myself that time is actually moving forward.

I stated that one of the efforts of this project was to rebuild memories that I had lost. With that as the measure, I would consider this to be a massively successful project. Even the pieces that are not in the data I was able to remember based on what was. Links of context, and the images that go along with it bringing back long forgotten moments.

8 - Conclusion + Future Expansions

It is difficult to judge how effective the installation is at showing the story hidden within my personal data, because it is a story about myself, and a life that I have lived. However through the development there were countless moments of remembering events, and people caused by examining and building the pieces. The way that I view the pieces is like looking through a portal in time, and seeing glimpses of what used to be.

Expanding this project into the future is something of great interest, as the collection of the data the installation is based on has continued, and will continue. The first addition would be including emotion analysis for all of the years featured, which could be done at little additional cost if there was more time allotted.

A future version of the installation may have data that fills in the story in ways that can’t be predicted. A more in depth analysis of photos would be an additional layer to consider in order to draw more depth from them as well. The project could reasonably be expanded into its own exhibition, moving through the stories told through a number of sets of data, from different messaging platforms to see how they differ in their portrayal of relationships, or contrasting with another person’s photos and messages, and seeing the overlaps.

Looking more carefully into the works, there are elements that differ from my own memories. There are people who appear that the data suggests are far more important than they were, and times that seem far more socially active than they were, because

the relationships kept through messaging don’t perfectly align with the in person relationships. The goal of attempting to portray a story hidden within data did indeed succeed in this way

However, the most interesting aspect of the installation would be to have an outside party attempt to put together the relationships, and attempt to understand who the people were. Though this will unfortunately have to wait until it is safe to hold a public display of the work. Until that time, the installation will serve as a time machine, and capsule, of this version of my own life.

8.1 - Artifacts

During the creation of each of my projects, I find myself finishing with a number of artifacts. Pieces that were generated either alongside the project, or because of it, but don’t fit in with the goal of the project. Due to the scale and nature of the project, there were a number of artifacts created from it, nearly enough to be their own exhibition. Each of these artifacts are topics that I want to explore further as an extension of this project, or potentially can be given life as entirely new projects.

References

1. Moss, Yelizaveta. “Essay Film.” *Cinema and Media Studies*, 2016. https://doi.org/10.1093/obo/9780199791286-0216. L. Courchesne, “Naked in Paradise”, courchel.net 2018.
2. Marker, Chris. *Sans Soleil*. France: Argos Film, 1983.
3. Morrow, Justin. “How Hitchcock Used Editing to Turn 'Rear Window' into a Masterpiece of Visual Storytelling.” No Film School. No Film School, November 4, 2019. https://nofilmschool.com/2014/07/alfred-hitchcock-editing-rear-window-kuleshov-effect.
4. Fung, Richard. *Sea in the Blood*. Canada: Fungus Productions, 2000.
5. Tyrell, Charlie. *My Dead Dad’s Porno Tapes*, Canada: NYT Op-Docs.
6. Manovich, Lev. “Representing Phenomena as Data” *Data*, 2019.
7.
8. Cram, Jacob. “Text Based Life.” *jacobcram.xyz*, 2017. jacobcram.xyz/textbasedlife.
9. Hare, Jason. “What Is Metadata and Why Is It as Important as the Data Itself?” Opendatasoft, August 25, 2016. https://www.opendatasoft.com/blog/2016/08/25/what-is-metadata-and-why-is-it-important-data.
10. Poitras, Laura, Glenn Greenwald, and Ewen MacAskill. “Edward Snowden: the Whistleblower behind the NSA Surveillance Revelations.” The Guardian. Guardian News and Media, June 11, 2013.

https://www.theguardian.com/world/2013/jun/09/edward-snowden-nsa-whistleblower-surveillance.

11. Carroll, David. “Co-Owning Our News Future.” Medium. Decentralize.Today, June 12, 2016. https://medium.com/decentralize-today/co-owning-our-news-future-6ba732df6207.
12. Channel4News. “The Great Hack's David Carroll Finally Sees His Cambridge Analytica Data”. YouTube. YouTube, 2020. https://www.youtube.com/watch?v=5Swqc2NjEXM..
13. Vincent, James. “Google 'Fixed' Its Racist Algorithm by Removing Gorillas from Its Image-Labeling Tech.” The Verge. The Verge, January 12, 2018. https://www.theverge.com/2018/1/12/16882408/google-racist-gorillas-photo-recognition-algorithm-ai..
14. Cram, Jacob. “UNMASK”, *jacobcram.xyz*. 2019.
15. Boesiger, Willy, and Lucy Nussbaum. Corbusier. Barcelona: Gustavo Gili, 1982.
16. Angulo, Luis. “Brutalist Architecture: Raw, Honest and Blunt.” ASM. ASM, April 21, 2016. http://www.asmscalemodels.com/blog/2016/4/21/brutalist-architecture-raw-honest-and-blunt.
17. “Minimalism Movement Overview.” The Art Story. Accessed May 28, 2021. https://www.theartstory.org/movement/minimalism/.
18. Reas, Casey, and Chandler McWilliams. Form + Code: in Design, Art, and Architecture. New York, NY: Princeton Architectural Press, 2010.
19. Wikipedia contributors, "Punctuated equilibrium," Wikipedia, The Free Encyclopedia,

https://en.wikipedia.org/w/index.php?title=Punctuated_equilibrium&oldid=1015575621 (accessed May 28, 2021).

20. Derivative, “Touchdesigner,” *Derivative*. Derivative.
21. Cram, Jacob. “Gallery.CRM”, *jacobcram.xyz*. 2020.
22. Wikipedia contributors, "Sonata," Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/w/index.php?title=Sonata&oldid=1020138519 (accessed May 28, 2021).
23. Robjohns, Hugh. Does Your Studio Need A Digital Master Clock?, June 1, 2021. https://www.soundonsound.com/techniques/does-your-studio-need-digital-master-clock.
24. Wikipedia contributors, "Great Red Spot," Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/w/index.php?title=Great_Red_Spot&oldid=1024649506 (accessed May 28, 2021).
25. “Emotion Analysis.” paralleldots. Accessed May 28, 2021. https://komprehend.io/emotion-analysis.
26. “Colours of Stars.” Douglas College Astronomy 1105. Douglas College Astronomy. Accessed May 28, 2021. https://pressbooks.bccampus.ca/astronomy1105/chapter/17-2-colors-of-stars/..
27. “Everything There Is to Know about Sentiment Analysis.” MonkeyLearn. Accessed May 28, 2021. https://monkeylearn.com/sentiment-analysis/.
28. “Render Baking.” Render Baking - Blender Manual, February 25, 2021. https://docs.blender.org/manual/en/latest/render/cycles/baking.html.
29. “A-Frame – Make WebVR.” A. Accessed May 28, 2021. https://aframe.io/.

30. “More than Just a Game: the Zen Power of Tetris | CBC Radio.” CBCnews. CBC/Radio Canada, August 30, 2019. https://www.cbc.ca/radio/tapestry/weekend-flow-1.5262870/more-than-just-a-game-the-zen-power-of-tetris-1.5263004.