


my sister in data

visualizing an online relationship

by Roxolyana (Sana) Shepko-Hamilton

A thesis exhibition presented to OCAD University in partial fulfillment of the requirements for the degree
of Master of Design in Digital Futures

Toronto Media Arts Centre
32 Lisgar Street., April 12, 13, 14
Toronto, Ontario, Canada
April 2019

 Roxolyana Shepko-Hamilton 2019



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Canada license. To see the license go to <https://creativecommons.org/licenses/by-nc-sa/2.5/ca/> or write to Creative Commons, 171 Second Street, Suite 300, San Francisco, California 94105, USA.

creative commons copyright notice

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material

Under the following conditions:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

Non-Commercial — You may not use the material for commercial purposes.

ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

With the understanding that:

Waiver — Any of the above conditions can be waived if you get permission from the copyright holder.

Public Domain — Where the work or any of its elements is in the public domain under applicable law, that status in no way affected by the license.

Other Rights — In no way are any of the following rights affected by the license:

- Your fair dealing or fair use rights, or other applicable copyright exceptions and limitations;
- The author's moral rights;
- Rights other persons may have either in the work itself or in how the work is used, such as publicity or privacy rights.

author's declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners. I authorize OCAD University to lend this thesis to other institutions or individuals for the purpose of scholarly research. I understand that my thesis may be made electronically available to the public. I further authorize OCAD University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Signature _____

Roxolyana (Sana) Shepko-Hamilton, OCAD University
My Sister in Data: Visualizing an online relationship
Master of Design in Digital Futures, 2019

abstract

This thesis proposes the exploration and visualization of social media data as a way of understanding human relationships in the age of the internet. The research includes the process of creating visualizations from a dataset composed of archival messages sent between my sister and I on Facebook Messenger between 2010 and 2018.

I employ autoethnographic methods to reflect on my exploration of the data, and use research through design methods in order to create interactive data visualizations. This exploration and creation process is guided by the following research questions:

1. In what ways can a visualization of my Facebook message data show how social media is used to maintain a relationship with my sister across changing distances and time?
2. How can data visualization help me learn more about the online connection between my sister and I?

acknowledgements

Firstly, I would like to thank my advisors, Ali Shamas Qadeer and Isabel Meirelles, for their insight, guidance, and wisdom. Their effort in helping to direct this project was invaluable, and I am incredibly grateful for the time, support and the positive attitude they provided. Thank you for being a lovely team to work with!

I am grateful to my colleagues in the Visual Analytics Lab, particularly Greice Mariano and Marcus A. Gordon, for providing support and helping to find solutions to technical issues.

I would also like to acknowledge and thank the Digital Futures Thesis professors who helped to inspire and shape the idea from its earliest stages to finished stages. This includes: Kate Hartman, Emma Westecott, Suzanne Stein, Adam Tindale, and David McIntosh.

A huge thank you to my fellow Digital Futures students for emotional support and for providing the space for mutual stress-ranting: you guys rock, and I am in awe of all of you.

I would like to express my deepest gratitude to my family, partner, and the friends who are basically my family. Thanks to my parents, for always believing in me even when I definitely didn't believe in myself; to my grandmother, for her strength and wisdom; to all my siblings, for always having my back (special shoutout to my brother Stefan Shepko for fielding my coding questions); to Orest and his parents, for always loving me and being my family away from home; to Daria and Morris Diakowsky for their warmth and encouragement; to Maria and Alina, for being my soul sisters.

Last but not least, thank you to my sister, Zoya Shepko, for being a constant presence in my life even when we are living hundreds of miles and an international border away. I love you so much. This thesis would not have been possible without your contribution.

table of contents

- Introduction** 1
 - Rationale 3
 - Scope 4
 - Theoretical Framework..... 4
 - Overview..... 5
- Literature review** 6
 - 1. Identity, relationships and sister bonds 7
 - 2. Technology as an Extension of Self 10
 - 3. Analysis of Social Media 12
 - 4. Privacy and Personal Data Collection..... 14
 - Conclusion..... 15
- Research methodology** 17
 - Data visualization methods 18
 - Autoethnographic methods 20
 - Participants..... 20
 - Ethical Considerations..... 22
 - Conclusion..... 23
- Process** 24
 - Prototype 1 25
 - Prototype 2 34
 - Prototype 3 44
- Conclusion**..... 56
 - Implications..... 57
 - Future work..... 59
- Bibliography** 61
- Appendices**..... 64
 - Appendix A - Screen Time and writing..... 64
 - Appendix B - Accessible charts..... 80

List of Figures

Figure 1. Data visualization steps.	18
Figure 6. Most repeated message content. Colour indicates sender.	26
Figure 7. Size of circle shows the amount that message content is repeated.	27
Figure 8. Clusters show frequency of message content, by year.	27
Figure 9. Sketch for network graph visualization.	29
Figure 10. Repeated message contents.	29
Figure 11. An image of the initial step of building the node dataset in Excel...	31
Figure 12. The link dataset, showing the source as "Zoya sent a photo..."	31
Figure 13. Message hover 1.	32
Figure 14. Message hover 2.	32
Figure 15. Network graph with prototyped interaction design.	33
Figure 16. Sketch 1.	35
Figure 17. Sketch 2.	35
Figure 18. Sketch 3.	36
Figure 19. Sketch 4.	36
Figure 20. Wireframe of the first view of Prototype 2 visualization.	38
Figure 21. User flow of Prototype 2.	40
Figure 22. Overlapping area graph created in D3.js	42
Figure 23. Data view.	42
Figure 24. Wireframe of proposed visualization for Prototype 3.	45
Figure 25. Tableau visualizations showing number of messages sent in each year	47
Figure 26. Comparison of graphs showing the number of messages sent by Zoya	48
Figure 27. Comparison of graphs showing the number of messages sent by Sana	49
Figure 47. Screenshot of the coded visualization from Prototype 3.	52
Figure 48. Close-up of the time axis annotation on the posters in my thesis exhibition.	53
Figure 49. A picture of the full thesis installation setup.	54
Figure 2. Nov. 4, 2018.	65
Figure 3. Dec. 2, 2018.	65
Figure 4. Jan. 6, 2019.	65
Figure 5. Feb. 3, 2019.	65
Figure 28. November 4, 2018.	68
Figure 29. November 11, 2018.	69
Figure 30. November 18, 2018.	70
Figure 31. November 25, 2018.	71
Figure 32. December 2, 2018.	71
Figure 33. December 9, 2018.	72
Figure 34. December 16, 2018.	73
Figure 35. December 23, 2018.	75
Figure 36. December 31, 2018.	75
Figure 37. January 6, 2019.	76
Figure 38. January 13, 2019.	77
Figure 39. January 20, 2019.	78
Figure 40. 1.	79
Figure 41. 2.	79
Figure 42. 3.	79
Figure 43. 4.	79
Figure 44. 5.	79
Figure 45. Black and white version of Figure 6 on page 28.	80
Figure 46. Black and white version of Figure 7 on page 29.	81
Figure 47. Black and white redesign of Figure 8 on page 29.	81
Figure 48. Black and white version of Figure 13 on page 35.	82
Figure 49. Black and white version of Figure 14 on page 35.	82
Figure 50. Black and white version of Figure 15 on page 35.	83
Figure 51. Black and white version of Figure 22 on page 44.	84

Figure 52. Black and white redesign of Figure 24 on page 47.	85
Figure 53. Black and white version of Figure 25 on page 49.	86
Figure 54. Black and white version of Figure 26 on page 50.	87
Figure 55. Black and white version of Figure 27 on page 51.	88
Figure 56. Black and white version of Figure 47 on page 54.	89

introduction

In 2017, a study surveying 1,500 Canadians found that 84% had a Facebook account (Gruzd et. al), and other research indicates the numbers of active social media users escalates every year. A Pew Research Center survey of United States adults shows a steady increase in the percentage of adults who use social platforms from 2012–2018 (Smith and Anderson). In January 2018, a report from We Are Social and Hootsuite (both of which are prevalent social media analysis companies) revealed that more than 3 billion people across the world use social media each month, with 9 out of 10 of those people using mobile devices to access social media platforms (Kemp). The popularity and ubiquity of digital media and social networking grows with each generation, and along with the growth in usage, so does the number and types of platforms that are popular amongst different groups of users. The variety of social media platforms contributes to a dispersion of the kind of content that is posted. According to comprehensive research on global social media usage, presented in *How The World Changed Social Media*, individuals locate “niches” of social relationships within different platforms, and behave according to the audience to whom they are presenting content (104). This adheres to theories proposed by Erving Goffman regarding the concept of ‘framing,’ which derives from his analysis of social activity (Miller et. al 103-104). The diversity of social media usage indicates a rich psychological landscape that can offer meaningful insights into the multiplicity of digital identity, as well as insight into the complex nature of negotiating relationships with other people across various social media platforms.

Although most social media can be downloaded for free on mobile devices or accessed easily in a computer browser, the usage of social media has a price. Edward Snowden's infamous leaks and the Cambridge Analytica scandal—two instances of the exploitation of personal internet data—highlight the fact that internet platforms are constantly tracking user interactions, messages, and behaviour. And yet, the immense quantity of information collected might be able to tell a story about people from the perspective of computational data, providing a unique opportunity to understand humanity from content that is in some ways self-authored and in other ways structured by the demands and desires of capitalist interest. Data can construct a profile of a person, separate from that person's physical self, that is created because of a medium that enables much of the social interactions of modern life. As Judith Donath writes,

“...a person's role in her social world... her web of social ties, whether numerous and diverse, or fewer but dense and close—is a major part of her identity... depicting a person's social network and the communication that flows through it can be meaningful, evocative...” (221). At the core of this thesis is an investigation of these digital footprints of social media interactions, and how they may represent the connections and relationship between two people; a relationship which can exist entirely in its own digital space, a world that is arguably just as real as the physical space which we occupy (Turkle, “Alone Together” 153).

The purpose of this thesis is to examine the overwhelming presence of social media in many people's lives, its impact on the maintenance of close relationships, and the data collection that occurs as a consequence of engaging with social media networks. *My Sister in Data* traces the digital relationship between my sister and I, across changing time and distance. The research examines our Facebook message history, downloaded from my personal Facebook user profile. The dataset starts in 2010 (a point when my sister and I had lived together our entire lives), progresses to when I moved away from home for the first time in 2017, and ends in 2018, after the first full year we had lived apart.

This research explores the following questions:

1. In what ways can a visualization of my Facebook message data show how social media is used to maintain a relationship with my sister across changing distances and time?
2. How can data visualization help me learn more about the online connection between my sister and I?

My Sister in Data includes the creation of a visualization using the archival dataset of private messages between my sister and I on Facebook over the course of 8 years. At the outset of this timeframe, my sister is 14 and I had just turned 16. The interactive visualizations created from this dataset are ultimately a reclamation of collected social media data, which includes our personal messages and interactions. Private messages, especially within Facebook, appear to be protected through an illusion of privacy. As Judith Donath writes, "...these actions are happening in a space that is not only public, in that many eyes can see it, but is also hyperpublic in that it can be seen for an extended time, in many contexts" (Donath 283). This research makes use of our chat history, thereby making a perfect example of Donath's points on the collection of content and the recontextualization of information. Chat messages are seen as

ephemeral, and yet this project makes the data archive hyperpublic, exposing seemingly fleeting dialogue and displaying it in a completely different context, because Facebook's storage of data allows it. This visualization intentionally exposes collected personal data to an audience as a way to reclaim the content that has been posted to an internet platform, where personal data is debatably not solely our own property anymore. In some ways, this is an empowerment of identity and relationships online, as well as an attempt to understand a rapidly changing social world and my own personal bond with my sister throughout time. In parallel to the creation of the visualizations is an autoethnographic process through which I reflect on exploring the archival dataset and the memories it evokes. This includes noting the emotional effects of moving away from my family and the ensuing consequence that the majority of my interactions with my sister are now maintained through social networking platforms. The process of visualizing archival Facebook message data allows me to observe our sisterhood from a broader, external view, while the autoethnographic process ties me to the internal emotional undercurrent that grounds this thesis.

Rationale

This thesis research is situated within the fields of communication theory, new media, and technological posthumanism. It focuses not only on the way that the internet affects human relationships, but also how we can understand human relationships in the age of the internet through the visualization of social media data. My interest in this research stems from the fact that the way that we receive and distribute content has transformed very quickly in the era of big data. Not only does this change have consequences on the way that "content" in the standard sense is understood, but also on the way that personal data becomes content that informs digital marketing trends and personalization of what appears to users on the web. Humanity's use of technology has changed our world drastically, and thus it should be acknowledged that each technological advancement sparks a fundamental transition in the way that we function as a society. There is a lot of existing research on the effects of digital technologies on academia and learning (Katherine Hayles' *How We Think*), the effects of social media on relationships and sense of self (Daniel Miller et. al's *How The World Changed Social Media* in addition to the majority of Sherry Turkle's writing) as well as the data collection enacted by social media platforms (Judith Donath's *The Social Machine*). However, this thesis proposes an integration of existing research with techniques of data visualization

and autoethnography in order to create prototypes out of my own Facebook dataset. In this way, I can present a case study of what the conversations between two sisters navigating young adulthood looks like in data collected by a social media platform, in the hopes that it helps to further the general understanding of what it is like to grow up using and communicating through social media; something that has affected my generation greatly, and will assumedly affect many generations after.

Scope

This research is intended to be a case study of relationships online, and only consists of two participants and our interactions on Facebook. As such, this thesis does not intend to generalize the experience of all people who use social media to maintain close relationships, but rather to provide an example of what a close relationship can look like through data. Creating a technical prototype, especially by using tools and coding languages that I am fairly unaccustomed to, takes time and iteration, and this takes away from allowing me to explore more than one person's dataset. In addition, the research does not necessarily address the effects of content personalization on social media platforms, which is ultimately the purpose of personal data collection; however, there are other works of research that focus on the consequences of filter bubbles and content personalization on the web, such as Ania Medrek's 2018 graduate thesis at OCAD University, *News by Association: Designing a way out of the echo chamber*.

Theoretical Framework

This work builds on writing by theorists such as Sherry Turkle and Katherine Hayles, whose work is concerned with technology's general societal effects as well as its specific impact on our relationship with others. More directly, *My Sister in Data* relates to Judith Donath's argument that in order for social media to be truly sociable, it is necessary to create thoughtful interfaces that allow a process of reflection on the ways that a beneficial social world can translate to a digital environment. Her book *The Social Machine* describes this argument and exemplifies data visualizations created from data collected on social platforms, including visualizations of conversations, network graphs, and online social landscapes. Donath's emphasis on applying thoughtfulness and reflection to social media use align with my own desire to discover how social media has played a role in the connection between my sister and I, and how it might help enhance my understanding of our digital sisterhood.

My Sister in Data is supported by two sociological research studies on sibling bonds. The first research study, conducted by Thomas Lee et. al, finds that high levels of communication in adult siblings is shown to be the case in instances when adult siblings live nearby, or in the case that the siblings are emotionally close. The second study by Katherine Davies supports the theory that children who are siblings construct individual identities through a process of comparison to their siblings, and that existing narratives of differences exist amongst siblings, reinforced by their parents; i.e., one sibling may be identified as troublesome, and the other as well-behaved. As the Facebook message data I am choosing to visualize is between my sister and I, these sociological studies help to frame my research and provide theories to describe certain insights that exploring the data uncovers.

In addition, a crucial aspect of this research is exploring the possibility of demonstrating the existence of a digital relationship by using data. The dataset downloaded from Facebook encompasses only the social interactions enacted through its platform, which obviously separates from physical social interaction; as Judith Donath puts it, "Online, there is no body; there is only information" (9). This naturally leans towards posthumanist theory, particularly technological posthumanism, which is a concept that seeks to reconsider what it is to be human in an age of advancing information technology.

Overview

This chapter provided an introduction to my thesis' background, intentions, and theoretical framework. The document's next chapter expands on the theories introduced in this section, and organizes the literature review according to four themes: *identity and relationships*, *technology as extension of self*, *social media*, and *privacy and personal data collection*. Next, the research methodology chapter, guided and supported by the insights from the literature review, explains my use of research through design and autoethnographic methods in application to my project. Finally, my process chapter outlines my research, results and reflection, categorized by the documentation of my first, second and third prototypes.

literature review

This chapter describes several areas of research and theory in order to present a comprehensive understanding of the complexity and interdisciplinary nature of the background of this research. This chapter is partitioned into various subsections of themes; these include *identity and relationships*, *technology as extension of self*, *social media*, and *privacy and personal data collection*.

Firstly, in the section “Identity, relationships and sister bonds,” I discuss theories of identity as proposed by Erving Goffman and Jenny L. Davis, who consider identity a type of performance. In this section, I also discuss research on digital relationships, guided by Erin Manning’s writing on engendering, a concept that describes touch as a way to create new matter that would not have existed outside of that connection. I apply the concept of engendering to research conducted on digital relationships by Anderson and Brown McCabe, who write about the opportunities that the internet creates to engender thriving online communities. I also discuss Judith Donath and Daniel Miller’s contrasting perspectives on internet relationships. “Identity, relationships and sister bonds” concludes with an examination of sociological research on the topic of connection between siblings, looking at Thomas Lee and Katherine Davies’ research studies. The inclusion of research on siblings is important in application to my thesis, which focuses on the case study of visualizing the message history between my sister and I. Thomas Lee researches the maintenance of communication between adult siblings, while Davies studies construction of narrative and identity among children who are siblings, which she finds is established through a process of comparing and contrasting oneself to one’s siblings; in other words, becoming what the other sibling is not.

The second section, “Technology as an extension of self,” highlights the importance of Sherry Turkle and Katherine Hayles’ writing to this work, and expands on the idea of posthumanism, which Katherine Hayles describes as the potential to augment human capabilities with technology and artificial intelligence. This section also mentions Marshall McLuhan’s influence on the idea of technology as extension of self. McLuhan’s work is based in the concept that any technology, or medium, transforms the way that the content it transmits is perceived. This indicates that technology has an influence on people’s understanding of information. Neil Postman builds on McLuhan’s ideas, and suggests that the medium

not only affects the way that content is perceived, but that technologies are built with an inherent bias and predisposition towards certain uses. Hector Huyke similarly echoes these sentiments, and posits that all technologies have become an extension of humanity, each with their own intentionality. The concept of technology as an extension of self is pertinent to my research, as I am studying the traces of a human connection between my sister and I that is wholly contained within the structure of Facebook's platform, which itself is a technology, enabled because of many other technologies.

"Technology as an extension of self" leads thematically into the third section, "Analysis of social media." This portion of the literature review primarily helps to outline what is considered social media. Definitions range from danah m. boyd's assertion that social media is web-based and socially connected, to Daniel Miller et. al's understanding that it is a platform which allows a user to move from a position of varying privacy levels and different sizes of social groups. The section also highlights the effects of social media on the user, citing its influence on self-perception and its negative impact on overall well-being. The negative impacts are contrasted by opinions presented by Daniel Miller, Wellman and Hampton, and Judith Donath, who value social media's ability to create virtual connections between people.

Finally, this literature review concludes with the section "Privacy and personal data collection," an important acknowledgment of the data collection enacted by large social media platforms—data collection that is ultimately rooted in the desire to make a profit from targeted advertising. This section highlights instances of personal data breaches and leaks, such as Edward Snowden's exposition of personal data collected by the United States government, and the Cambridge Analytica scandal, which involved the utilization of Facebook user's data in order to inform Donald Trump's 2016 election campaign. This section also researches various social media platforms' user privacy statements, including Snapchat's, Facebook's, and Instagram's privacy policies.

1. Identity, relationships and sister bonds

Core to my thesis is an investigation of how technology transforms people's ability to interact with others. In order to understand this, it is important to explore concepts of identity and relationship in order to apply this lens to my research. In addition to this, the prototypes draw from a dataset of a conversation between two sisters on Facebook messenger, and thus it is beneficial to my research to include studies on sibling and sister relationships.

a. Identity

Erving Goffman's *Presentation of Self in Everyday Life* is grounded in the concept that people can and will change their behavior in order to control how others perceive them, leading them to perform their identities as if in front of an audience. Goffman introduces the term *dramaturgical analysis* as a way to understand this identity performance. Jenny L. Davis builds on Goffman's and others' writing about the self in her article "Triangulating the Self: Identity Processes in a Connected Era," and applies theoretical concepts to the contemporary digital environment. She argues that people "strike a presentational balance between ideal and authentic" by *self-triangulation*, a method of "presenting a coherent image in multiple arenas and through multiple media" which is partly enabled by a fluidity between the physical and digital (500). Davis writes about the notion of networked logic, which "refers to networked individuals' seamless incorporation of these multiple media into performative practices" (513). In turn, networked logic affects the performance of self to others—or in Davis' words, a negotiation of identity in online and offline environments (500). Similar to Goffman, this suggests that a presentation of self is preferably controlled, in all social aspects. These ideas about identity align with Sherry Turkle's assertions in *Alone Together* that online performance is integral to the construction of social media profiles and avatars. "There is nothing more deliberate than the painstaking work of constructing a profile or having a conversation on instant messenger in which one composes and recomposes one's thoughts" (276). Performance, therefore, is a crucial aspect of understanding and applying concepts of identity both offline and online.

b. Relationships online

Relationship is defined by Oxford Dictionary as "the way in which two or more people or things are connected, or the state of being connected." In this way, I see relationships as an extension of identity, in which two performative entities combine to create a new connection in relation to each other. Erin Manning, in her book *Politics of Touch*, discusses the notion of engendering—"The engendering body is a body not only in relation but of relation. Relation is activity, intensity, movement toward. Engendering is an event through which emergent bodies take form" (104). Manning argues that through the act of touch it is possible to create or engender new matter that would not have existed outside of that connection. This posits that a connection between people allows for an emergence of something that could not exist were it not for the connection.

Manning's perspectives on relations between others is interesting to consider when applied to digital situations. In a time when so much interaction, or "touch," occurs online, the internet becomes a hub of thriving engenderings. The research paper "A Coconstructed World: Adolescent Self-Socialization on the Internet" by Anderson and Brown McCabe discuss the impact the internet has on young people's construction of identity and the importance that online connection can have—"Broadened connections gave informants more opportunities to meet new people, some of whom they would not have met otherwise in the offline world. Some informants were simply broadening the opportunity to establish new relationships..." (245). Conversational dialogue within internet platforms can reflect an amalgamation of connection. However, Judith Donath remarks astutely that in many cases, digital spaces don't offer a sense of presence, or permanent identity: "Online text conversations are very different. The audience is invisible. There is no shared space, no gestures, and no visible identity. Even in synchronous text chats, where all the participants are online at the same time, there is relatively little sense of presence. You are visibly present so long as you are typing, but quickly fade away if you are inactive..." (248). On the other hand, Daniel Miller writes that it is misleading to believe that online relationships must necessarily be compared to their real-world counterparts; simply, the internet gives people another way in which to connect to each other, "much the same way that everyone treats the landline telephone, never described today as a separate 'online/ on-phone' facet of life" (100). Wellman and Hampton second this notion, and believe that computer-mediated connections are as authentic and meaningful as those in the 'real-world' (651). In this way, digital technology can be seen as a facilitator of posthuman connection.

c. Sibling and sister bonds

A digital dialogue between two sisters is at the core of this work, and research conducted by Thomas Lee et. al shows that sister-sister kinship pairs have the highest degrees of contact compared to other types of kinship bonds (Lee 483). This research study also discusses predictors of communication between siblings, assigning emotional closeness as one of these factors: "Emotional closeness includes a sense of shared experience, trust, concern, and enjoyment of the relationship. Research has shown that among the reasons why kin stay in touch are common interest and admired personality characteristics..." (Lee 433) as well as a feeling of responsibility for the family member (Lee 433). Lee et. al's research also demonstrates the importance of proximity in the general maintenance of kinship relationships, citing that

respondents who lived closer to their family had more interaction, but importantly noting that this interaction appears to be obligatory. "Living close made people feel obligated to be in more contact with one another than they would have been by choice" (Lee 438). By drawing from Lee et. al's research, it can be suggested that the contact maintained between my sister and I is not out of obligation—since in the past year we have not been geographically proximate—but out of our emotional bond.

On the other hand, Katherine Davies' research focuses on the construction of narratives amongst children who are siblings. Her study "Siblings, Stories and the Self: The Sociological Significance of Young People's Sibling Relationships" is a qualitative exploration of the relatedness and situatedness of identity amongst siblings. Davies' research reveals that there are existing narratives of comparisons amongst siblings; "Many respondents compared themselves to their siblings without being prompted to do so, but what is striking about those who were specifically asked to make comparisons is that most required little or no time to consider their response, suggesting an existing narrative surrounding sibling similarities and differences which participants were able to readily draw upon" (Davies 684). This study reveals that the interaction and hierarchy of siblings within a family structure affects the way that children create their identity, which is often constructed out of relating how different or similar they are to their siblings; "The way Sadia describes one sister as 'compromising' the other less helpful and well behaved sisters is illuminating in that it suggests an understanding that, when taken together, the characteristics of the three sisters form a well-balanced whole. Sadia is the 'naughty one' because her sisters are not and this identity is of course also constructed within the complex gendered dynamics of these relationships" (Davies 685). Siblinghood, therefore, is a factor in the construction of individual identity. As demonstrated through Davies' research, children who are siblings develop an identity in relation to their siblings, often through a process of comparison. What one sibling lacks, the other provides. In addition, Lee et. al's research shows that being a sibling as an adult may lead to regular communicative contact in the case of close proximity, or emotional closeness.

2. Technology as an Extension of Self

Social media is incredibly influential and has been integrated into many aspects of the majority of people's lives. Given its prevalence of use in addition to the other digital technologies employed for the purpose of communication, it can be argued that mobile devices, social media profiles, and other technologies all

function as an extension of self. As Turkle writes in *Alone Together*, "These days, being connected depends not on our distance from each other but from available communications technology. Most of the time, we carry that technology with us" (155). The fact that these technologies allow people to expand out of natural physical limitations, such as spatial distance, indicates that technology extends human capacities. Posthumanist theory, as outlined by Katherine Hayles in her book *How We Became Posthuman*, is a discussion of this phenomena, primarily the idea that human potential can be augmented and entwined with artificial intelligence and other technologies. Hayles, however, critiques the posthumanist privileging of information and disregard for material self. She writes that her ideal vision of a posthuman world "embraces the possibilities of information technologies without being seduced by fantasies of unlimited power and disembodied immortality, [one that] recognizes and celebrates finitude as a condition of human being..." (5).

An early and pivotal contributor to the concept of technology as an extension of humanity is Marshall McLuhan. His book *Understanding Media: The Extensions of Man* was conceived and written during the 1960s, at an earlier stage of technological development than our contemporary era. However, many of his ideas still apply today, as he primarily wrote about various medias and their impact on how content (or "the message," to use McLuhan's terms) is interpreted. McLuhan famously argues that "...the medium is the message. This is merely to say that the personal and social consequences of any medium—that is, of any extension of ourselves—result from the new scale that is introduced into our affairs by each extension of ourselves, or by any new technology" (23). He also warns that it is "only too typical that the 'content' of any medium blinds us to the character of the medium" (24). Likewise, Neil Postman expands on McLuhan's thoughts in his book *Amusing Ourselves to Death* by acknowledging that technologies contain inherent bias, and a predisposition toward certain uses; thereby making it impossible for technologies to remain neutral (84). This connects to McLuhan's reference to the "character" of the medium, which Postman and McLuhan both agree is inherent. Similarly, in his paper "Toward an Ethics of Technologies as Prosthesis" Hector Huyke argues that all technologies have become an extension of humanity; technology is created with an intentionality or goal, which becomes a prosthesis.

More directly related to digital technology, Katherine Hayles' book *How We Think* is based on the idea that the human thought process is formed and defined by the type of media consumed by an individual. Hayles points out that the transition from print to digital is something that will affect the way our brains function, even going so far as to suggest that the more someone works with the digital, the more their brain will become "enmeshed within larger networks that extend beyond the desktop computer into the environment" (3). She coins the term 'technogenesis,' or in other words, "the idea that humans and technics have coevolved together" (10). Hayles' concept of technogenesis consequentially implies that our use of technology, and in particular the internet, can function as an extension of self.

3. Analysis of Social Media

a. Comparing definitions of social media

Variations on definitions of social media include danah m. boyd's assertion that it is any "web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system" (210). This can be compared to Daniel Miller et al.'s definition of social media as media which is scalably sociable, meaning that "each of these platforms corresponds to a position of greater or lesser privacy and smaller or larger groups" which can be adjusted according to the user's needs; "Groups and platforms may overlap, but mostly we find platforms become associated with specific genres of communication which people see as appropriate for the group engaged with that particular platform" (5-6). Miller's observation aligns with Turkle's thoughts in her paper "Computational Technologies and Images of the Self," where she describes a user's activities on the computer as a display of various windows through which they can achieve different tasks, which she refers to as a summary of the user's distributed presence on the computer. "In practice, windows have become a potent metaphor for thinking about the self as a multiple, distributed system. According to this metaphor, the self is no longer simply playing different roles in different settings... The life practice of windows is of a distributed self that exists in many worlds and plays many roles at the same time" (1101). From this it is possible to conclude that internet activity, and therefore social media, prioritizes various types of performances and content depending on the context and audience in question, and that the user is often engaging in multiple online roles at a time.

b. Effects of the use of social media on the user

Digital technology and social media's influence is widespread and consequently has an impact on the way that people interact and live their lives. Social platforms fluctuate in popularity and usage, which is a phenomenon that is noted by academic writers within this field; additionally, the way in which a social media site or mobile application is used leads to a set of unique and sometimes unpredictable consequences, depending on the situation. For example, *The Guardian* published an article in August 2018 about 'Snapchat Dysmorphia,' a name for the increasing requests for plastic surgery in order to reflect the appearance one might have in a Snapchat AR filter (Wolfson), while Facebook has notoriously gone through multiple scandals regarding their misuse of user's personal data; New York Times writes that "trust in the social network [Facebook] has sunk, while its pell-mell growth has slowed" (Frenkel et. al).

Recent studies show that social networking platforms have a negative impact on overall well-being; "...the tricky thing about social media is that while we are using it, we get the impression that we are engaging in meaningful social interaction. Our results suggest that the nature and quality of this sort of connection is no substitute for the real world interaction we need for a healthy life" (Shakya et. al). Several writers argue that digital technology's effects also include a dissociation from the information we read online, in addition to a disconnection from other people. As Turkle writes in *Alone Together*, "Networked, we are together, but so lessened are our expectations of each other that we can feel utterly alone. And there is the risk that we come to see others as objects to be accessed—and only for the parts we find useful, comforting, or amusing" (154). Geert Lovink, in "After the Social Media Hype," writes about the "attention fatigue" that occurs because of the immense amount of information that is encountered online; "We can read as many facts as we like, but if we try to add them up, they refuse to become a system. We struggle to keep track of all the information that approaches us, making it hard for most info bits to be properly digested... When we are constantly exposed to real-time interactive media, we develop attention fatigue and a poor sense of time." Turkle and Lovink both indicate the problems that arise out of an environment marked by immense amounts of information and possibilities for connection, leading to a decline of productive consumption. On the other hand, there are others who counteract this cautious perspective of social media use. Miller calls this critical perspective as an "anxiety about the new

technology" (102), and Wellman and Hampton refer to it as "dystopians who fear that computer-mediated ties are inauthentic or less meaningful..." (651). Scholars who comment on the more positive outcomes of social networking sites mention how "computer networks enhance connectivity, and help people form and maintain ties over long and short distances" (Wellman and Hampton 650), an opinion which is echoed in Judith Donath's book, *The Social Machine*—"We are at the beginning of a revolution in human communication. The online world has the potential to connect each individual to millions, if not billions, of others. This is an unprecedented scale for forming communities, for meeting people, and for working together" (6). It is undeniable that technology has advanced to the point where it is extremely easy to contact someone even if they are across far removed geographically, which allows for the formation of a global community. These varying realities of social media's effects should be considered when analyzing its impact, for not only is it a tool for creating connections which would be impossible otherwise, but through this enablement constructs its own set of limitations and problems.

4. Privacy and Personal Data Collection

Finally, it is important to reflect on another side effect of social media in today's world: the immense amount of data created within and collected by online social platforms. The circumstances of social media data collection are ultimately what allows the formation of my thesis to take place, therefore making it crucial to acknowledge its practice. Edward Snowden's exposition of material from the United States' National Security Agency in 2013 is just one example of a public realization that the data produced by people when using their mobile devices or the internet can be tracked and collected by governments and corporations. In *The Guardian's* interview with Snowden, he claims that "the value of the internet, along with basic privacy, is being rapidly destroyed by ubiquitous surveillance" (Greenwald et. al). Similarly, the Cambridge Analytica scandal in 2016 revealed Facebook user's data was collected by an application within Facebook's platform, titled This is Your Digital Life. This app collected data that was used by Cambridge Analytica, a data consulting firm which worked with Donald Trump's election team and the Brexit campaign in the UK (Cadwalladr). In the summer of 2018, "Zuckerberg Responses To Judiciary Committee Questions For The Record," as part of Facebook's Congressional hearings in response to the scandal, Facebook stated that that the company collects "(1) data about things people do and share (and who they connect with) on our services, (2) data about the devices people use to access our services,

and (3) data we receive from partners, including the websites and apps that use our business tools... We use data... to obtain these interests and to personalize every aspect of our services..." (15). Of course, Facebook is not the only social platform collecting its user's data; Instagram's privacy policy states that it collects data on the content provided by users, their interactions, messages, social network, device data, IP address, among other things ("Data Policy," *Instagram Help Center*); Snapchat's data policy is similar, although it suggests that they don't collect data on the actual images or messages shared through their platform ("Privacy Policy," *Snap Inc.*). In an article written for *The Atlantic*, the author recounts her experiences downloading her Facebook dataset—"The data download was a time capsule of sorts, a rare record of time spent digitally... [it] is a consolation prize offered to those of us on the losing end of surveillance capitalism. The folder underscored some of social media's most unappealing qualities: the distortion of a natural, human experience of time, and an insistence on never quite letting things go." She describes the process of sifting through her self-created data as giving her an opportunity to view herself "through the eyes of Facebook's partners, researchers, advertisers, and algorithms, in an act of reverse surveillance" (Wiener). Humanity's development of tools for connection evolve within a capitalistic climate, making it unclear whose best interests are considered. Platforms claim data collection is primarily for enhancing user experience; and yet, it also conveniently enables targeted advertising. In Bruce Schneier's *Data and Goliath*, he discusses data as a "by-product of high-tech socialization" (1), and that the amount of data created per day increases exponentially; in 2015 (the year the book was published) seventy-six exabytes of data traveled across the the Internet (5). Nick Srnicek's *Platform Capitalism* defines platforms as a business that provides the hardware and software for others to operate in. He discusses how internet platforms have impacted the economy, with data as one way "to maintain economic growth and vitality in the face of a sluggish production sector... The platform has emerged as a new business model, capable of extracting and controlling immense amounts of data..." (Srnicek). Ultimately, the allowances to broaden connections that the internet grants to its users often comes at the cost of sacrificing immense amounts of personal data.

Conclusion

The four themes highlighted in this literature review are all important to support and ground my research, which aims to understand the digital relationship that I maintain with my sister on Facebook through the

creative design practices of data visualization. On a preliminary level, investigating concepts of performative identity as well as online relationships and sibling bonds are crucial to the core of this research, and will help to understand the ways in which I might be able to interpret our interactions through the Facebook platform. Secondly, applying the concept of technology as an extension of self and the theory of posthumanism enriches the research. It allows the interpretation of the Facebook message data to be perceived as a demonstration of the relationship between my sister and I despite the fact that it is enacted through a digital platform, and not associated to our physical selves. By grounding this thesis in the concept of technology as an extension of self, I can assert that our Facebook message data is affirmation of a posthuman connection. Third, an understanding of how to define social media and its larger consequences on users is an important contextualization for this project, and I note that I have personally experienced many of the negative associations referenced with social media use. And finally, the section "Privacy and personal data collection" highlights the social media data collection makes this research possible, and emphasizes the importance of critically assessing the many pitfalls that personal data collection can lead to.

In the next chapter, I discuss the methodological framework used to conduct the research, which includes methods from research through design and autoethnography.

research methodology

This thesis researches the connection I maintain with my sister on the internet through the process of creating data visualizations out of our Facebook message data. This project employs a research through design (RtD) methodology, thereby validating the act of design as a unique form of research. RtD is used in tandem with methods from autoethnography, allowing the particular case of personal data used in this project to be contextualized in the reflective methods of an autoethnographic approach.

Research through design is defined as an approach that incorporates methods and techniques from design practice as a process of inquiry; i.e., a methodology which acknowledges the method of creation as a form of research (Zimmerman). This framework is helpful to my thesis interest as a whole because much of the research is rooted in the discoveries that creating the prototype unearths; therefore, answering my research questions only begins to be possible after using my Facebook dataset to visualize and analyze results. As put by Groeme Sullivan in *Research Acts in Art Practice*, "...research acts can also be interactive and reflexive whereby imaginative insight is constructed from a creative and critical practice. Oftentimes what is known can limit the possibility of what is not and this requires a creative act to see things from a new view" (Sullivan). RtD, therefore, is in itself a reflexive practice that allows the researcher to consider and learn from the act of making. Iteration and reflexivity is invaluable in this particular research endeavor, and helps to highlight creative practice as a form of research—in this case, using data visualization as a tool to understanding large datasets, and the insights that those datasets may convey.

On the other hand, autoethnography contextualizes and places an importance on the personal data at the core of the research. The personalization of this work is unavoidable, due to the intimate nature of the dataset itself. Autoethnography is also a way to use personal accounts as a form of research into larger cultural experiences (Ellis et. al), and this allows the thesis to situate itself as an example of people's experience with the state of current digital technology and social media. Morgan Brigg, in the article "Autoethnographic International Relations," writes that the perspective of the author/researcher should always be acknowledged; "The voyage cannot be erased, and neither can the

framing, the fading, the restoration work. To erase the author is to erase potentially important insights: it leaves us with less knowledge rather than more" (780).

Autoethnography and research through design, then, both support the idea that process (be it of creation or discovery) is in itself a meaningful form of research. These methodological traditions are connected in the idea that results of research cannot be separated from the context from which they originate, and in a similar vein, in designing a data visualization out of the Facebook message history between my sister and I, it is impossible to separate my own personal experience or the process of creation from this research.

Data visualization methods

This thesis employs methods outlined by Ben Fry in his book *Visualizing Data* (2008). The series of steps include: acquiring the dataset, parsing, filtering, data mining, and representing the data. However, I will note that the organization of some steps is slightly antiquated; for example, Fry places adding interaction in the visualization as the final step of the process, which in the age of digital-native web graphics should be considered earlier in the stages of creating a data visualization.

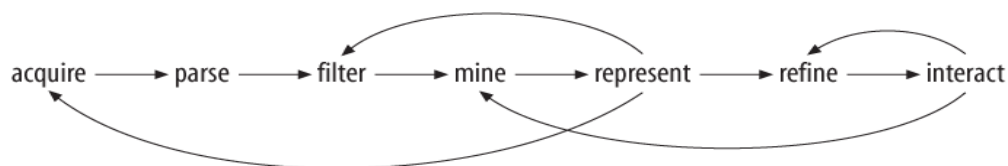


Figure 1. Data visualization steps.

Fry, Ben. "Computational Information Design." PhD dissertation at Massachusetts Institute of Technology, 2004, p. 30.

1. Acquiring the dataset

In the case of this research I focused on my personal dataset from Facebook. This is due to the fact that it has historically and contemporaneously been the primary method of communication between my sister and I. In order to access the dataset, I navigated to my user settings while logged into my Facebook account, and then went to the tab "Your Facebook Information." From here, I selected "Download Your Information," which allowed me to select what aspects of my Facebook data I wanted to include in the dataset (this encompasses everything from my posts, search history, photos, comments, to even records of my payment history on Facebook). I eventually would download two copies of my Facebook data, at two different points during the research; first in August 2018, and the second time in January 2019. In

both instances, I opted to download all possible data available to me from Facebook. I identified the folder which contained my and my sister's message history. In addition to the message .JSON file (which includes all of the content of the message text), this folder also included the files, photos, videos, audio, and gifs that we have sent to each other. For the purpose of the visualization, I concentrated on the folder containing our message history.

2. Parsing the data and creating a structure

This step involves understanding the data and organizing it into categories. Since the complete dataset downloaded from Facebook was already set up in a very organized folder structure, and the data files themselves were set up as .JSON (this stands for JavaScript Object Notation; files set up as .JSON are already ordered into hierarchical structures), reorganizing the dataset was not a necessity unless it was required for the specifications of the visualization type I was trying to build out through code.

3. Filtering data

In this case, this involved a removal of any other information that did not pertain to the interactions and messages between my sister and I. Other user's data, or content that referred to other people, such as mentioning a name of a mutual friend in our message data, was removed from our message data through a basic search and replace functionality enabled through most code text editors (in this particular case I was using a combination of Sublime, Brackets and Atom IDE). After searching a mentioned person's identifying name, I would replace their name with "[name]" in order to anonymize any reference to other people. Filtering data in order to protect other people's privacy was also a necessary process in the photos sent between us. Our dataset includes a lot of screenshots, and therefore it was necessary to block out the names of other individuals in the images we sent to each other.

4. Data mining to find patterns in the dataset

Data mining involves an iterative process that includes visualization and experimentation to understand the patterns in the datasets. In my case, I primarily used Tableau for data mining, a quick visualization software that doesn't require any programming. I expand on this step later in the [Prototype 1](#) section of my process chapter.

5. Representing the data through visualization

The experimentation of the data mining phase helps to define extractable insights and aspects of the data that can help to build a narrative. After data mining, it is possible to pick the approach that will be used to create the visualization. This includes choosing an appropriate visual metaphor, as well as chart type and design of the overall experience of the visualization. In the case of my research, I chose different visualization approaches in my first, second and third prototypes, according to my changing understanding of the information that I would like to convey through the visualization.

6. Refinement of the design

This step adheres to the iterative prototyping technique that is common amongst many creative practices. Once a visual approach has been defined, refining and tweaking the visual approach helps to ensure that the visualization is understandable and communicable.

7. Adding interactive elements to visualization

According to Ben Fry's outline of visualization methods, the interactive elements are the last step of the process; however, this step in the methods has been considered throughout the entire process of my research.

The data visualization techniques listed above are important to help explain the process of creating data visualization. It is also important to mention that in this research, Ben Fry's visualization methods are used in the context of a variety of different visualization and prototyping tools, including Excel, Tableau, and D3.js to name only a couple.

Autoethnographic methods

Autoethnography is an approach that uses personal experience as a way to understand a larger social phenomenon. In my case, the methods used as part of my autoethnographic approach are the following:

1. Documenting my personal experience being a participant of and audience to social media culture.
2. Reflective writing in tandem with the exploration of the Facebook message data between my sister and I, in order to uncover my reactions and experiences to the archival data.

Participants

The participants involved in this research are my sister and I. I chose to focus on this particular relationship because historically (as well as contemporaneously) she is one of the most consistent social bonds that I've had in my life. This history is reflected in the dataset from Facebook—our message

archive dates back to a message from Zoya in 2010, making this one of the most chronologically expansive datasets in my Facebook message history. This, in addition to the fact that we are siblings who care for each other, made this particular dataset meaningful to my thesis, which undertakes identifying and finding insights from a personal relationship in social media data.

In order to situate me and Zoya within the context of our family, I will provide a brief summary. I was born in November 1994, and Zoya was born 15 months later. We have two younger brothers, and a younger cousin that grew up with us after her mother's death. Our parents met in graduate school in our hometown, and have been married for 25 years. My grandmother on my mother's side was born in Ukraine, and escaped from our home country with her brother and my great-grandmother during World War II. Her father was an artist and was executed during the Great Purge, a campaign enacted by the Soviet Union to eliminate so-called 'counter-revolutionaries.' My grandmother now lives across the street from our family farm. My American grandmother on my father's side was diagnosed with Alzheimer's, and has been living with my family for around four years with my mother as her primary caretaker. Zoya and I, along with our other siblings, were homeschooled throughout our childhoods and grew up very conjointly. Much of our socialization was dependent on our family, and we often played outside together until I reached adolescence, at which point I began reading a lot of books and separating myself from my younger siblings. We were encouraged to take music lessons when we were growing up; we both began playing violin around the age of two, and piano around age eleven. Later, Zoya abandoned violin as an instrument and focused her attention on learning the cello, while I stuck with violin. Music was a very important part of our lives while growing up, and it led to the creation of our band, which originally started out as a family ensemble and has now morphed to include other musicians. Zoya and I frequently discuss our band both in person and online. My sister and I both attended the same university where our parents met, and therefore we all lived at home for almost all of our lives. Finally, 2017 was the first year we lived away from each other, when I moved to Toronto to attend graduate school at OCAD University.

Technology use in our family was encouraged primarily by our father. He made us an "Apple family"; i.e., a family that predominantly uses Apple products. I remember our first computer from the early to mid-2000s, that barely connected to the internet and was mostly handy for basic games. Later, he gave us our first iPods for Christmas, and I received my first laptop computer in 2011 as a hand-me-

down. It also did not connect to the internet, but I used it for journaling and writing. We had a family iMac desktop computer that everyone used. Around 2011, my siblings and I had iPod touches, which are essentially iPhones except they aren't able to be used as a phone or connect to data. Around 2013 onwards, we used our own personal computers more often than we used mobile devices, until Zoya got her first iPhone in 2016. I got an iPhone two years later in February 2018.

I acknowledge that as a participant of the online conversation between my sister and I, my perspective regarding the results of the research is subjective. However, being a participant of our Facebook conversation also allows me to have a deeper understanding of the context that our conversations take place in; through the process of reading through all the messages again and through seeing the results of the data in the visualization, I can garner snippets of the ways that data from Facebook can illustrate moments from our lives more holistically.

Ethical Considerations

It is important to highlight that the dataset used to power the prototype is downloaded from my own Facebook account, focusing on the private conversation data between me and my sister. This exposes information from both me and my sister that would otherwise not be revealed to a public audience. Because of this, consent from my sister needed to be ongoing. An application to the Research Ethics Board was necessary, and helped to formulate protective measures for my sister's rights in the development of this work. One way that I attempted to mitigate this issue was to offer the option of setting up bi-weekly meetings, for her to check in on my progress on the prototype and to let me know if there's anything that she's uncomfortable with sharing. Another option I offered to her was to have a backend administrative interface that would allow her to manually remove message data, or to automatically remove all of the data if at any point she wanted to wholly remove the consent for me to use her messages in this project. And finally, I offered her the option to have access to the raw dataset I was intending to use for the prototype and to go through it with me to let me know if there should be any messages removed from this dataset. This being said, my sister is a busy person, and the options that involved her regular involvement were not particularly encouraged; regardless, it was important that she was informed of her rights and the opportunities available to her to be as involved in the development of the project as she desired.

Conclusion

In this chapter, I discuss my methodological framework of research through design and autoethnography and its application to my thesis. I use Zimmerman's definition of research through design as a methodology that acknowledges the method of creation as a form of research, as well as Groeme Sullivan's assertion that creative research practice is interactive and reflexive. Ellis et. al's definition of autoethnography as a methodology of using personal accounts as a form of research into a larger cultural experience helps to guide my framing of the research as a case study of the relationship between my sister and I in order to provide insights into a larger cultural phenomenon.

Also in this chapter, I outline the data visualization techniques that I employ in my research, deriving from Ben Fry's *Visualizing Data*, and highlight my use of reflection and journaling as an autoethnographic method.

In the following chapter, I recount my research process, which is separated into four subsections; the creation of my first, second and third prototypes.

process

This project is guided by methods from research through design and autoethnography. My research through design process employed iterative prototyping as a method to create multiple visualization sketches using a variety of tools, including Microsoft Excel, Tableau, D3.js, and Adobe XD. This allowed me to explore the Facebook chat data between me and my sister, which in turn aided my understanding of the parameters that were available in the dataset that could be used in a more refined visualization. The iterative process of prototyping the data visualizations was supported by autoethnographic written documentation that contextualized my own lived experiences. I initially incorporated a discussion of my iPhone's weekly Screen Time usage as a technique to discuss my relationship with technology. However, this framed my autoethnographic documentation to be primarily focused on my personal relationship with technology, rather than my lived experience understanding my technology-mediated relationship with my sister. In addition, the time-frame of the Screen Time tracking coincided very little with the Facebook message dataset. This ultimately did not assist the progression of my thesis, and therefore the Screen Time documentation is removed from the process chapter and relocated in Appendix A. I attempted to redirect my autoethnographic observations to focus more on my reflection of how examining the dataset helps to understand and describe the connection I have with my sister. This marks a significant change in my autoethnographic documentation. From the beginning of November 2018 to the end of February 2019, my documentation (listed in detail in the [appendix](#) section) is styled around the tracking of my Screen Time usage. Following February, my autoethnographic documentation is less chronological and more intertwined with my research through design documentation. My second stab at autoethnographic writing, subsequently, is part of the prototyping process, and can be found synthesized within the Prototype 2 and Prototype 3 sections. This chapter takes the methods mentioned in the previous methodology chapter, and applies them to the research process. The chapter is broken up into the following sections: documentation of the first prototype, documentation of the second prototype, documentation of the third prototype, and my thoughts regarding future steps.

Prototype 1

The first prototype followed a process of creating small data sketches to find ways in which I could visualize the data, and then choosing a chart type to develop as an interactive visualization. The eventual outcome of this prototype phase was a zoomable network graph showing the most commonly sent message text content, and the messages that directly preceded the common message. The nodes were shown according to sender, signified by the color of the node in the network. In the following subsections I describe my process in detail, which culminates in a reflection of the outcome of this prototype.

Data sketches

In my research methodology chapter, I outline the [data visualization steps](#) taken from Ben Fry's *Visualizing Data* that I use within my research through design methodology. At the outset of beginning the first prototyping phase, I had downloaded the dataset from Facebook and filtered out mention of other people in the data, including censoring any mention of other people in the photos sent between my sister and I. My next step was to understand the message data file; what information it contained, what parameters could highlight meaningful insights, and what patterns I might be able to find in the data. In order to begin experimenting and understanding the dataset, I chose to use Tableau, a data visualization software that enables creating visualizations out of datasets within the Tableau interface without requiring programming knowledge. I regularly use it in my data visualization practice, and find that it helps to understand the dataset in the early stages, which helps guide the design of the visualization later on. It also can be used to filter and edit datasets by exporting the dataset from its interface, which in the development of Prototype came to be helpful.

Using Tableau's bar chart visualization, I discovered the message text content that was most frequently sent between me and my sister.

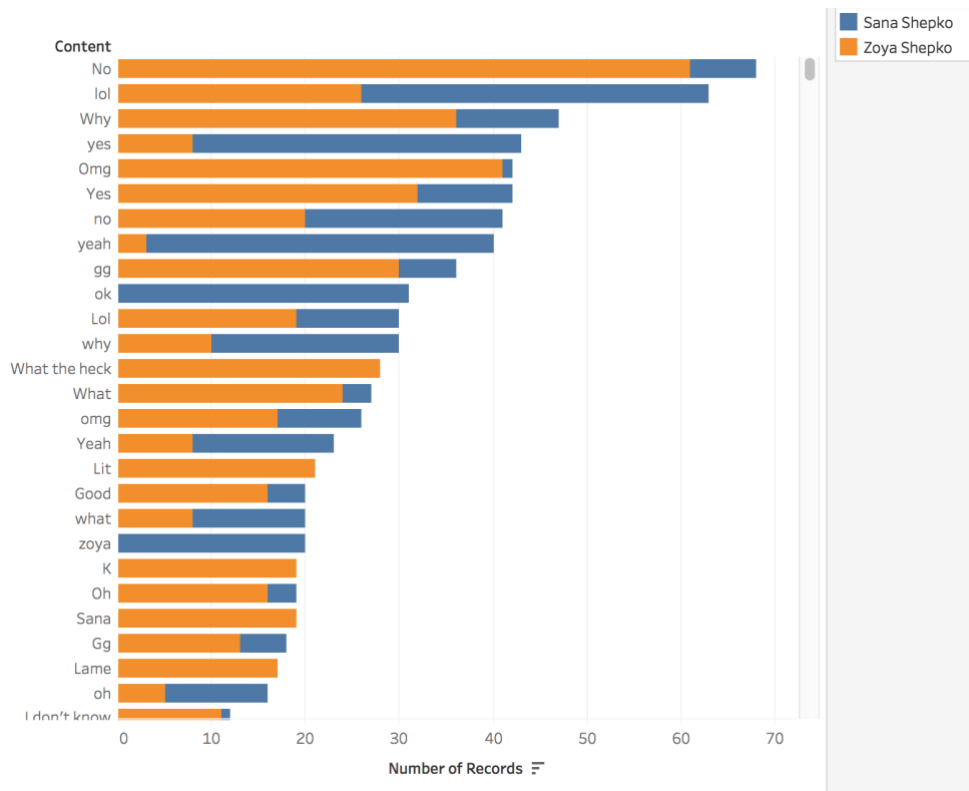


Figure 6. Most repeated message content. Colour indicates sender.

I became interested in these high number of records in repeated message content, because I felt that it could show trends in the specific language that was common between my sister and I, as well as typifying our message interactions without necessarily overwhelming a viewer with all of the data at once. Since I personally participated in this message dataset, I had a sense of what the highest frequency message content might be; but it was still very interesting to see the differences between what my sister sends and what I send.

I continued to develop other ways to show the message content frequency in Tableau. My next iteration was a bubble chart that displayed messages with a repeated content frequency of at least 2 instances or more. Each repeated message content bubble was associated with a specific color to indicate the sender of the message. The highest frequency message content was located in the center of the bubble chart, while message content with lower frequency was located towards the outer edges of the visualization.

2016. Another aspect to note is that in 2015, I began to take trips away from home more often, as that was the year that I began a long-distance relationship. This may help to explain the spike in messages sent back and forth after 2014, as the communication between my sister and I was often mediated through Facebook during the times that I was away from home. However, I cannot positively conclude this because other factors may have contributed to this increase in messages, such as a general increase in usage of technology, a cultural shift to being more digitally connected, etc. Finally, it is important to note that while the visualization shown in Figure 8 did incorporate progression of messages through time on a yearly level, I was also interested in understanding how messages were distributed across smaller time scales, such as number of messages over months rather than complete years. Therefore, the design of Figure 8 was abandoned due to its limitation in showing smaller time scales.

In order to move forward in the visualization progress of Prototype 1, I decided to change my toolset, and move from creating quick data sketches in Tableau to developing an interactive visualization in D3.js—an open-source JavaScript library that is primarily used to create dynamic interactive data visualizations through the implementation of SVG. The transition from Tableau to D3.js was assisted by Tableau's ability to export .CSV files from visualizations that had been edited within its interface. Having created the visualization that was centred around the highest frequency message content within Tableau, I exported a .CSV of the edited dataset. My next step required thinking of ways to create a prototype that would reflect the most frequently repeated message content that had been derived out of the Tableau iterations. In addition to this, a possible addition to this visualization could be the links between frequent message content, as well as identifying trigger messages and how they link to the frequent message contents. What relationships do frequent contents of messages have with other messages? What might visualizing this information reveal about the data?

For the first prototype, I decided that I would try to create an interactive, zoomable network graph. Network graphs, as defined by Google's Fusion Tables, are a type of visualization that show relationships between entities, which are displayed as nodes with links to show the relationships between them. This method of visualizing data can reveal shared commonalities between entities. In my plan for the network graph, central nodes of network clusters are frequently repeated contents of a message. The trigger

message is shown as an outer node which connects to the central node of the cluster, the response to the trigger message.

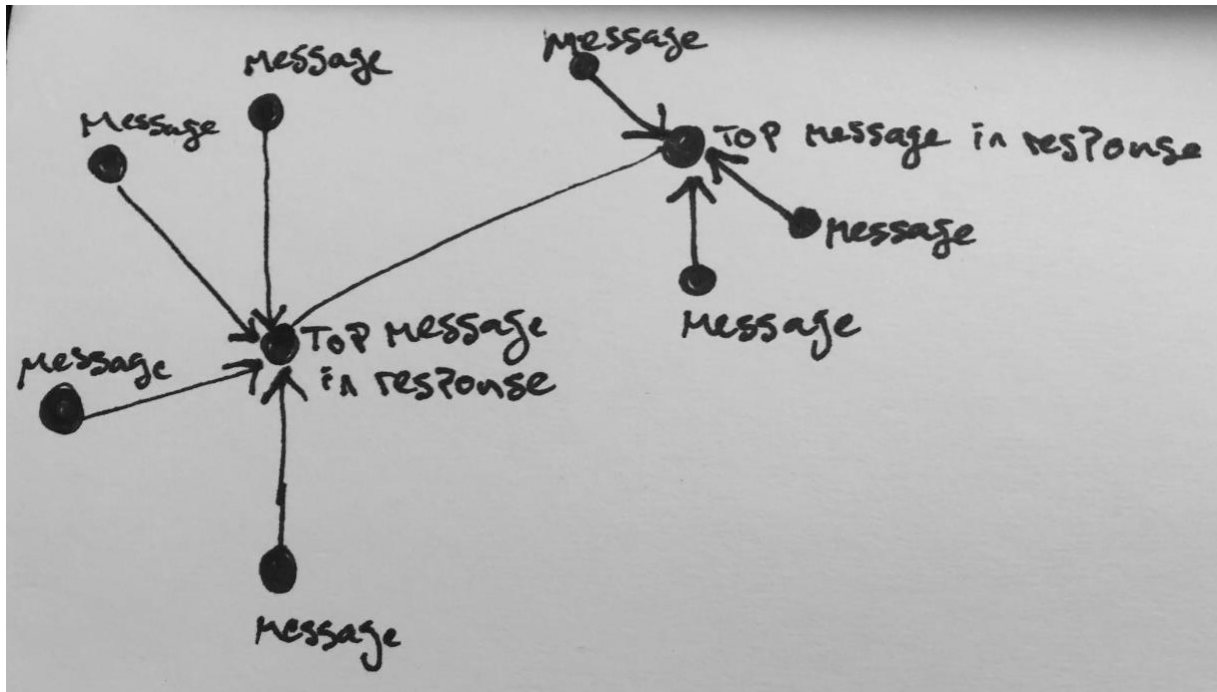


Figure 9. Sketch for network graph visualization

Having decided that I was going to attempt to incorporate the messages with the most frequently repeated content (messages such as 'No,' 'lol' and 'omg') as well as the messages that triggered the response, I had to revise the existing .CSV that I had downloaded from Tableau, which only included the data on the highest frequency message content.

TopMessages		
Content	Sender Name	Number of Records
Zoya sent a photo.	Zoya Shepko	106
Omg	Zoya Shepko	99
No	Zoya Shepko	93
Zoya sent 30 photos.	Zoya Shepko	90
lol	Sana Shepko	61
yeah	Sana Shepko	56
Yes	Zoya Shepko	48
gg	Zoya Shepko	48
Why	Zoya Shepko	44
What the heck	Zoya Shepko	42

Figure 10. Repeated message contents.

To create the revised dataset, I used the .CSV exported from Tableau as my starting template. In this .CSV, I would select a repeated message, and then search the repeated message in the data file containing all unique messages to identify each instance it appeared in the complete dataset. This allowed me to determine the trigger messages that had been sent in each instance directly before the repeated message. For example, the message content that is most repeated in the entire archive of my messages with my sister is when she sends a photo. This is documented in the dataset as “Zoya sent a photo.” I selected the contents of this repeated message, i.e., “Zoya sent a photo,” and searched it in the complete dataset of all unique messages. Every instance that it appears in the data, I selected the messages directly preceding it, the ‘trigger’ messages, and included the data from those trigger messages in my revised dataset. This process was repeated for the other most frequently repeated messages, until my revised dataset was completed enough to display a comprehensive enough visualization.

A network graph consists of connections between nodes via links. It requires a distinction of the node data and link data, which are usually separated as two different .CSV files or as two objects in a JSON file. It is crucial to include an identification number or name for each node in the node data; this enables the link data to then have an origin (the ID of the first node) and a target (the ID of the node to which it connects). The dataset I had been building needed to be revised again so that I could be able to visualize the network graph, since I did not have the node and link data separated. However, I had been building my dataset with an ID assigned to each type of message content, so it was easier to create the node and link datasets from this point.

id	Sender	content	daytime	count
14543	Zoya Shepko	Zoya sent a photo.	multiple	106
575	Zoya Shepko	I only have [name] until 7:30	12/13/2018 12:11:22 PM	1
635	Sana Shepko	lol lots of balloon photos	12/3/2018 8:19:33 PM	1
641	Sana Shepko	so many blurry photos lmao	12/3/2018 8:14:35 PM	1
788	Sana Shepko	Grad school will fucking kill	11/22/2018 9:13:14 PM	1
1039	Zoya Shepko	G	multiple	12
1214	Zoya Shepko	Rip	multiple	32
1237	Zoya Shepko	The only Christmas related t	10/29/2018 8:43:01 PM	1
1239	Sana Shepko	Can't even call	10/29/2018 8:08:54 PM	1
1271	Sana Shepko	Wanna send them to me	10/29/2018 5:37:06 PM	1
1318	Zoya Shepko	No	multiple	93
1327	Sana Shepko	Ok zoy	10/19/2018 11:31:52 PM	1
1387	Zoya Shepko	δ□□□	multiple	2
1460	Sana Shepko	i don't care that's very bad	10/10/2018 5:21:53 PM	1
1887	Zoya Shepko	lol	multiple	33
1982	Sana Shepko	ldk	multiple	11

Figure 11. An image of the initial step of building the node dataset in Excel, showing the top message “Zoya sent a photo” and the following rows as the trigger messages for the instances where “Zoya sent a photo” appears in the full dataset.

source	target	Year	sourcetime	targettime
14543	575	2018	12/13/2018 12:13:53 PM	12/13/2018 12:11:22 PM
14543	635	2018	12/3/2018 8:20:18 PM	12/3/2018 8:19:33 PM
14543	641	2018	12/3/2018 8:15:39 PM	12/3/2018 8:14:35 PM
14543	788	2018	11/23/2018 11:47:58 PM	11/22/2018 9:13:14 PM
14543	1039	2018	11/8/2018 11:45:53 AM	11/8/2018 11:44:56 AM
14543	1214	2018	10/31/2018 11:27:23 AM	10/31/2018 11:15:17 AM
14543	1237	2018	10/29/2018 9:05:16 PM	10/29/2018 8:43:01 PM
14543	1239	2018	10/29/2018 8:42:43 PM	10/29/2018 8:08:54 PM
14543	1271	2018	10/29/2018 8:02:21 PM	10/29/2018 5:37:06 PM
14543	1318	2018	10/22/2018 5:43:51 PM	10/22/2018 5:42:47 PM
14543	1327	2018	10/20/2018 12:04:18 PM	10/19/2018 11:31:52 PM
14543	1387	2018	10/10/2018 5:38:28 PM	10/10/2018 5:32:25 PM
14543	1460	2018	10/10/2018 5:22:06 PM	10/10/2018 5:21:53 PM
14543	1887	2018	9/29/2018 8:10:44 PM	9/29/2018 8:09:41 PM
14543	1982	2018	9/29/2018 3:34:06 PM	9/28/2018 6:35:47 PM
14543	2267	2018	9/19/2018 6:56:00 PM	9/19/2018 6:54:13 PM
14543	2276	2018	9/19/2018 6:52:38 PM	9/19/2018 6:52:02 PM
14543	2283	2018	9/19/2018 6:37:32 PM	9/19/2018 5:55:06 PM
14543	3315	2018	8/19/2018 4:23:25 PM	8/19/2018 4:16:19 PM
14543	4435	2018	7/25/2018 12:49:33 PM	7/25/2018 8:49:57 AM

Figure 12. The link dataset, showing the source as “Zoya sent a photo,” (ID 14543) and the targets as the trigger messages.

In creating this prototype, I followed several examples of network graphs, looking at Mike Bostock's work as well as others' work on bl.ocks.org, which is a viewer for sharing code that has been posted on GitHub Gist. In addition to visualizing the data as a network graph, I also wanted to have a tooltip appear when hovering over a node and to have the capability of making the graph zoomable. Because of this, I had to combine several bits of example code from bl.ocks.org in order to create the prototype that I was trying to achieve (authors of these bl.ocks examples included Mike Bostock, Caged, and Tom Roth). Some challenges ensued; I began building my dataset as two .CSV files, one for the node data and one for the link data, but because I was attempting to connect several bits of code together there were issues implementing this method of using the datasets, and I had to abandon the two separate .CSV files and instead build my node and link dataset in a JSON file.

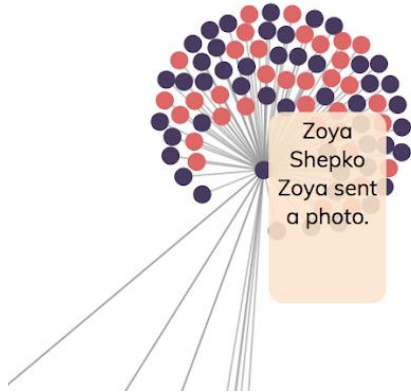


Figure 13. Message hover 1.

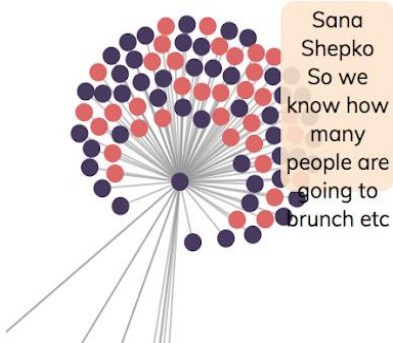


Figure 14. Message hover 2.

In creating the first prototype, I was mindful of assigning a specific color to each sender. It was also important that the contents of the message was displayed in a tooltip, although there were issues of readability in some instances. When thinking of the ideal design of this version of the prototype, I also was considerate of what the final design of the website might look like. The branding, therefore, was reminiscent of the branding of Facebook's messenger app itself, in this way contextualizing the data and the platform within which it was created.

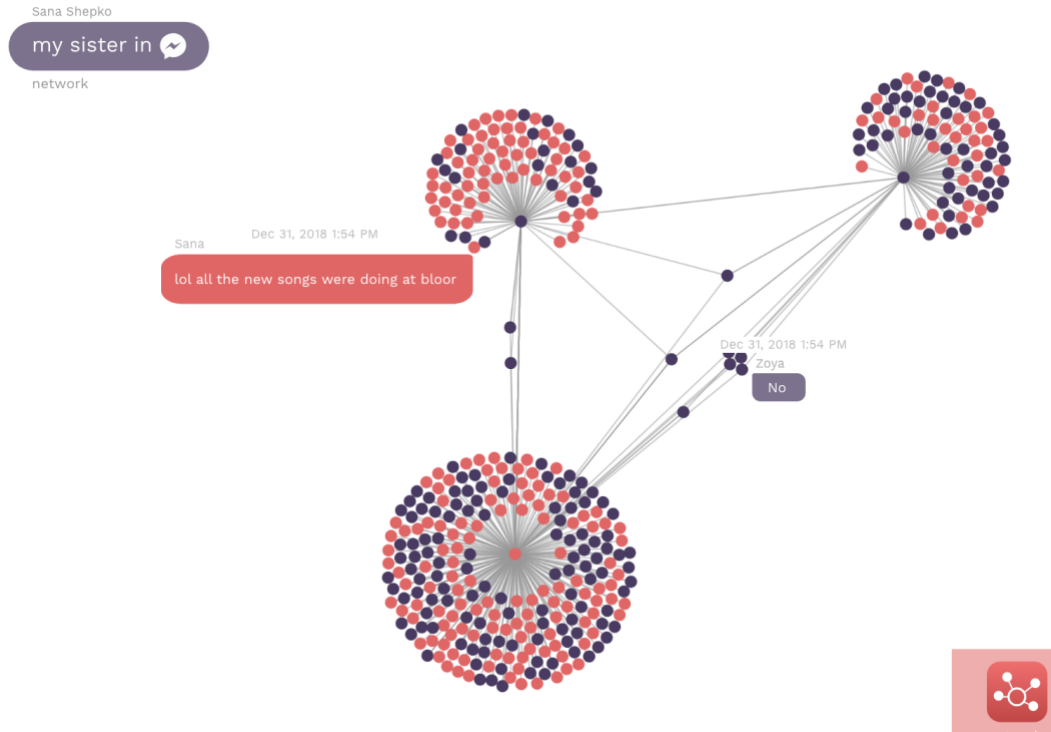


Figure 15. Network graph with prototyped interaction design.

Reflection on Prototype 1

It should be clarified that the method of using network graphs in this prototype is untraditional. As mentioned previously, they are used to show relationships between entities. In this case, I use the central node as an amalgamation of all the instances that that message had been in sent in response to the trigger messages, or the nodes surrounding the central node. This use of a network graph attempts to show cause and effect, rather than relationship, which ultimately may not be the best use of a network graph. In addition, an issue with this prototype is that it doesn't succinctly communicate any level of intimacy or hint at the context of a sister-sister dialogue spanning the course of several years. This prototype is a very birds-eye view of the dataset, and it requires a viewer to hover every individual node to get a sense of the conversation going on; however, this effort on the part of the viewer would still not necessarily lead to an overarching understanding of the content since the nodes are not organized in any chronological, progressive fashion. Instead, they are linked as triggers and response.

In reviewing this prototype with my advisors Ali Qadeer and Isabel Meirelles, we discussed the importance of using the research questions to guide the visualization. This prototype, although functional, did not lend itself to really answering how Facebook data can show a relationship between two people. It

was very much a technical exercise and perhaps a self-assurance that as a first-time D3.js user, I would be able to utilize this JavaScript library to create interactive visualizations. However, moving forward the aim was to work on designing a prototype that could directly address my research questions and contextualize the message data in a meaningful way. My advisors and I also discussed the importance of determining what my main goal was for the visualization. Outlining the main goal would help to build a narrative that could help to define the next steps for prototype 2.

Prototype 2

As I moved forward in understanding the goals of the research and what I hoped to achieve in the creation of this visualization, I realized that essentially, my ultimate aim is to use the Facebook dataset as a way to understand the online relationship between my sister and I. This includes identifying changes over time and distance, as outlined in my first research question, and on a more general level, using visualization of our Facebook message data to reveal insights about the online connection between us. The time frame of the message data includes changes in our lives in terms of where are we living in relation to each other, and this is also an important factor that adds to the meaning I would like to communicate to a viewer of the visualization. Does the dataset indicate changes in the way we communicate when we are farther away from each other? Do we use the internet as a substitute for our face-to-face interactions, and how might the dataset show this? What type of things do we communicate to each other using the internet, and are there changes in our communication that can be indicated over time?

Early ideas and sketches

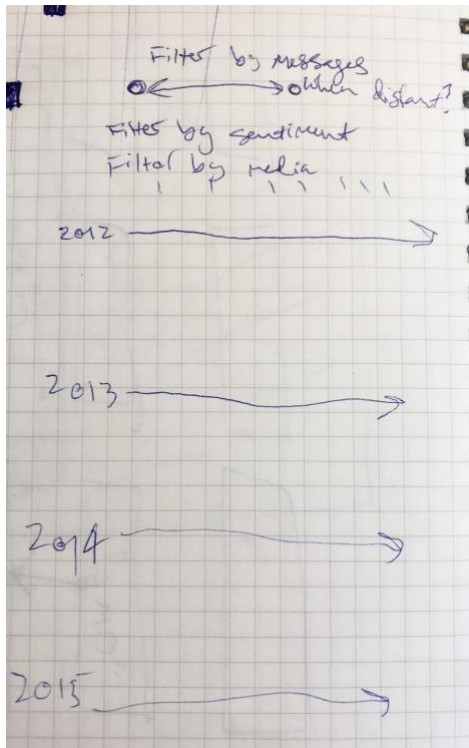


Figure 16 Sketch 1.

Through this active consideration of the research interests that lay at the heart of this work, I moved away from the network graph and towards a more linear chart, which would assist me in comparing changes over time, which is one of the aspects of my first research question. Another factor of my research question is identifying how data changes over distance, which also would be more easily identifiable if I were to use time as an axis, because then I would be able to recount during what timeframes we were close to each other in proximity and at what times we were geographically distant. My initial thought was to show each year with the contents of that year's messages scrolling horizontally, while the years increase as the user scrolls vertically. I also

considered using distance between my sister and I as a filter which could be applied to the graph. This initial sketch, while not quite achieving what I was trying get at, led me to other iterations.

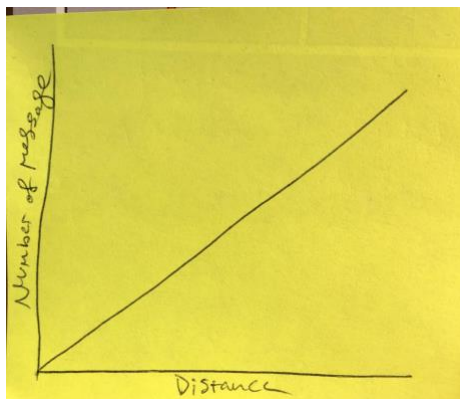


Figure 17. Sketch 2.

Next I considered how I might use the y- and x-axis as a starting point to explore my first research question: *In what ways can a visualization of my Facebook message data show how social media is used to maintain a relationship with my sister across changing distances and time?* I then decided to sketch an even simpler chart by directly comparing the number of messages to the distance between my

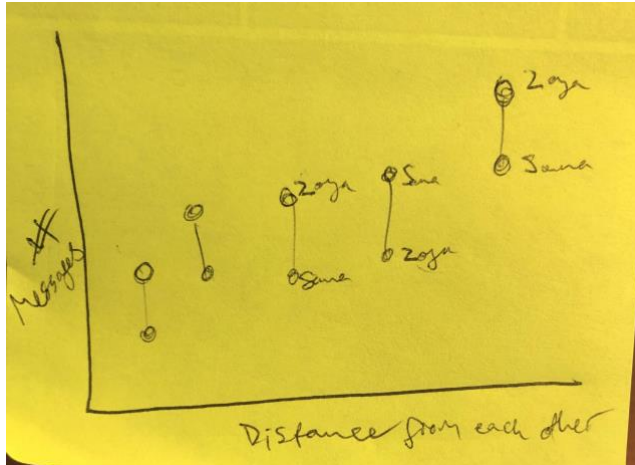


Figure 18. Sketch 3.

sister and I (this can be seen in the rough sketch above). Using distance and the number of messages as axes was getting closer to answering my research question, but it was still missing a way to show that this is a dialogue between two participants. Because of this, I felt it would be important to show that this is a comparison of the messages sent between two people, over time. Therefore, I started moving

towards a chart type that would allow me to show the messages sent from me and my sister in a comparable way.

Refining sketches

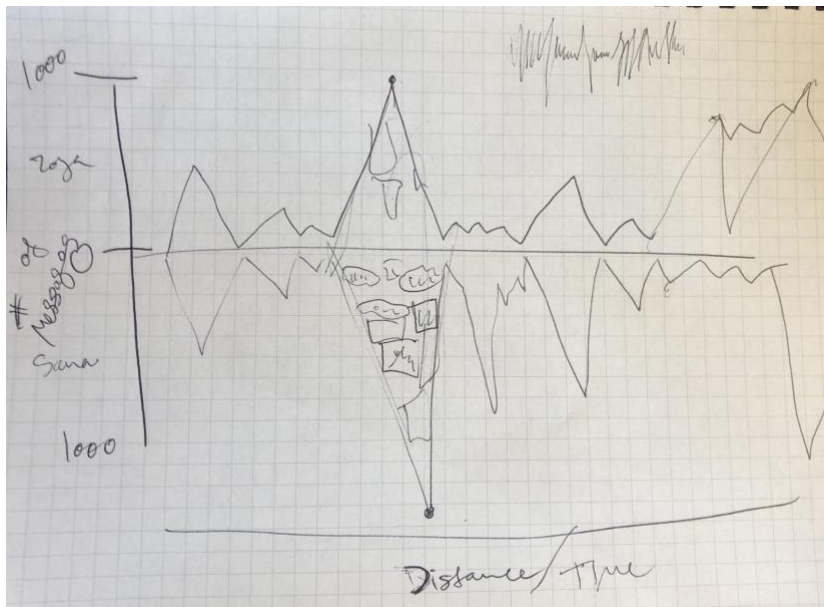


Figure 19. Sketch 4.

In the earlier sketches of prototype 2, the number of messages (the y-axis) represented all messages sent regardless of who the sender was. In these sketches the number of messages is a sum of all messages sent. However, it felt important for me to separate the messages sent by me and by Zoya to show what similarities or differences may exist. In "Siblings, Stories and the Self" Katherine Davies finds that siblings derive a sense of identity through the process of comparison to one another (692), and

therefore comparison was an important functionality for me to consider when designing the visualization. This separation and display of difference also helped to indicate the dialogue that occurred at that point in time or distance, depending on what the x-axis was. It occurred to me that it might be possible to include the content sent by me and Zoya separately, by splitting the axis that represents “number of messages” in half. Thus, the 0 point lies in the middle of the y-axis, and the maximum value of messages sent lies on the top and bottom of the y-axis. Messages sent by Zoya would be represented through half of the y-axis, and messages sent by me would be represented by the other half of the y-axis. This thought was inspired by visualizations of sound waves. Sound waves, if taken as a metaphor for the process of visualizing prototype 2, can imply the sound of a conversation or dialogue, which is what the dataset itself is composed of.

Ideally, I would have used distance as its own axis, but I realized that I have no way of using distance since it doesn't actually exist as a parameter in the Facebook message dataset, and I would have to personally annotate each message that was sent by me and my sister with my recollection of the distance we were apart from each other. Considering the limited scope of this project's lifespan, it would take too much valuable time to go through every record of our more than 14,000 messages and annotate the approximate distance away from each other for the time that message was sent.

In addition, I also wanted to determine if it might be possible to display the actual contents of the messages that were included in the graph. I decided that this might be possible using two area charts instead of two line graphs, and to populate the inside of the area charts with the contents of the messages sent by me and Zoya during that point in time.

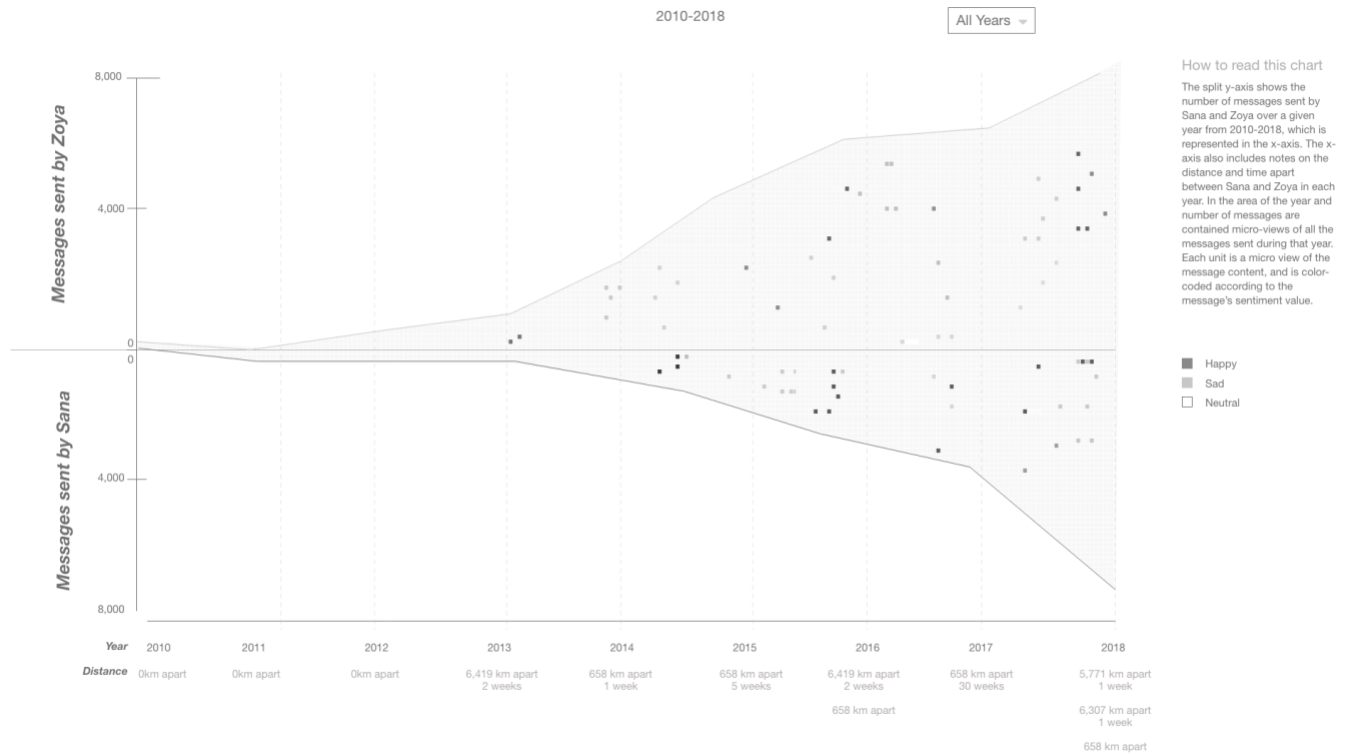


Figure 20. Wireframe of the first view of Prototype 2 visualization.

As I progressed with the idea of having a split y-axis displaying two area charts that represented the messages sent by me and Zoya over time, I thought about ways that I could include additional meaningful information in the graph. Displaying the contents of the messages, even at a very small size, could hint at what that message might contain. This small view of the message content could show whether the messages we are sending at that time are text, photos, gifs, etc, and could bring more meaning to the visualization regarding the nature of the conversation during that time. In addition, displaying the message content as a unit per each message also means that I could associate a color tint to the message unit. This color could represent the emotional value of that message; for example, blue tint on a unit would mean that this is a message that is sad in emotional content, and yellow tint on a unit would mean this is a message that is happy in emotional content. Considering that the amount of messages sent by either me or Zoya during the span of a year is at most around 7,000 messages, this means that 7,000 of these very small units colored by emotion could show an overall picture of what the emotional landscape was of that year's dialogue. It is important to note that at this time I have not determined the precise color scheme for the sentiment value or the methods of which the sentiment values will be

assigned on each message; however, if I continue to develop this prototype, I will be mindful of the ways in which sentiment value and colors are assigned to each message.

I still felt that indicating the distance from each other would be important to include in the graph. As mentioned previously, distance is not a parameter that is included in the dataset I downloaded from Facebook, and thus would require me to annotate the distance on each message that was sent based on my recollection of where we were at that point in time. Because of the possibility of inaccuracy as well as the amount of work this would require, it was not possible to use distance from each other as an axis. However, using time as an axis, I decided I could add text commentary showing the approximate distance that we were away from each other during that point in time.

User flow and interactivity

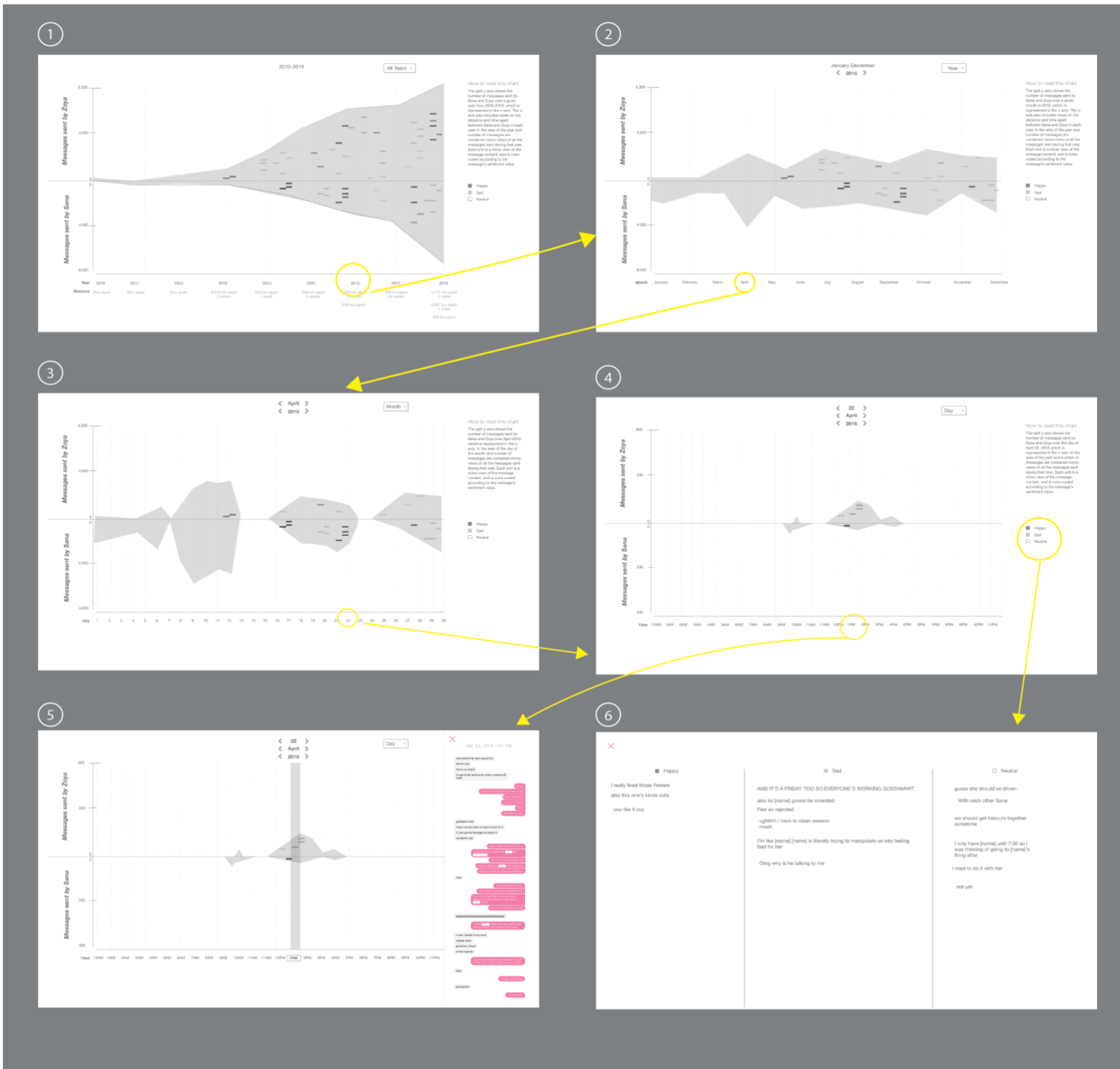


Figure 21. User flow of Prototype 2.

The interactivity of this prototype would allow a user to go increasingly more granular in time. In the above user flow, the first view of the visualization shows the time axis as starting at 2010-2018, all the years that

this dataset spans. If the user were to click on one of the years in the time axis, or in the section of the area graph that aligns with a particular year, they would be taken to a view of the graph that shows only that year; for example, a user clicks on 2016 along the x-axis, and navigates to a monthly breakdown of 2016, spanning from January 2016 to December 2016. If the user were to click on a month, for example, April, the view will show the days of the month of April 2016 and the message data that was sent during that month. If the user were to select a particular day in that month, then the view shifts to an hourly breakdown of that day, and finally, if the user selects an hour in that day, then a side panel opens up showing the conversation log of that hour. The color legend, which is located to the right side of every chart view in this prototype, explains how each message unit is assigned an emotion, and what emotions are represented by what colors. If the user were to click on the color legend they will navigate to another view of the data, where all the messages included in the dataset are organized by emotional categories.

Coding Prototype 2

As someone who is much more comfortable in design, one of the most terrifying moments is moving into code development. I knew, at this point in the prototyping phase, that I wanted to show a mirrored area graph along a split axis; and I knew that ideally I'd like to represent each message as its own unit in the visualization.

This being said, I'd like to strongly emphasize that implementing D3.js to create web-based visualizations is not my strength. I am strongly reliant on basing my code on existing examples. Because it is an open-source JavaScript library, luckily there are kind souls who post helpful code snippets to work off of; however, since the design of the visualization of Prototype 2 is rather unique, I wasn't able to display all aspects of the design that I was planning to when applying it in development. However, I was able to achieve creating an overlapped area graph showing messages sent over time.

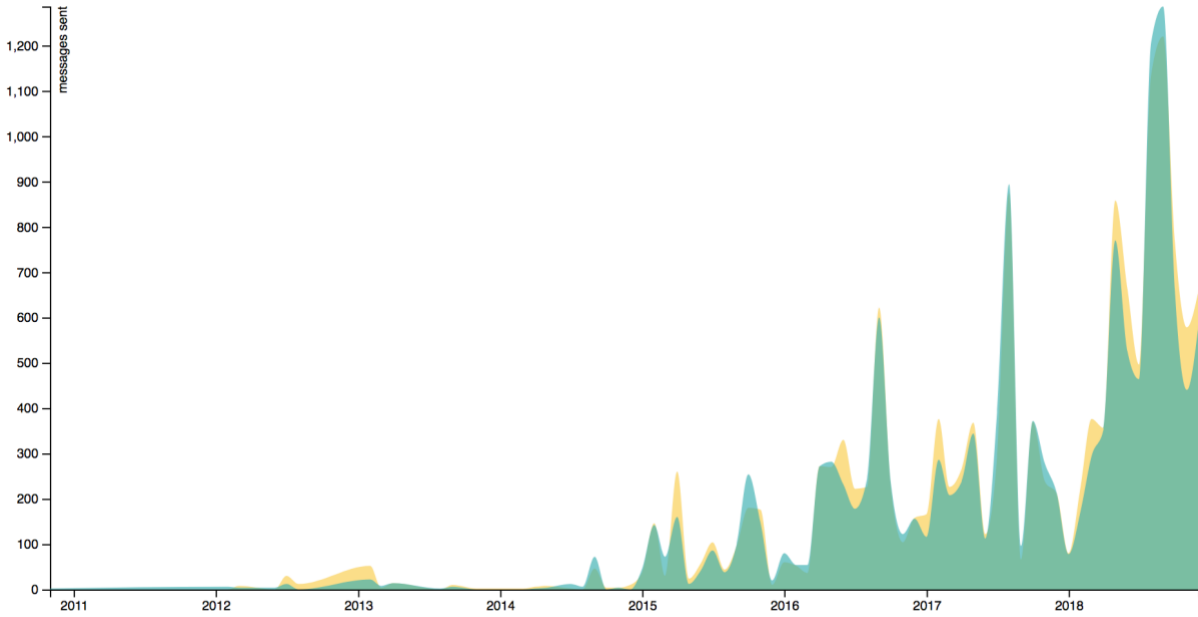


Figure 22. Overlapping area graph created in D3.js, showing number of messages sent over time in years. Green indicates messages sent from Zoya, yellow indicates messages sent from Sana.

There are several things to note about the process of making this overlapping area graph in D3.js. Firstly, in order to create this, I had to export a vastly stripped down version of the original .JSON data file out of Tableau as a .CSV file, which only included the number of messages separated by sender in each month of the 8 years represented on the x-axis. This leads to some limitations, especially considering the importance of being able to display and represent each message as its own unit, rather than representing a sum of messages sent in a timeframe. Displaying individual messages as units would allow me to

date	Sana	Zoya
2010/11	0	2
2012/02	0	6
2012/03	8	4
2012/06	0	4
2012/07	30	12
2012/08	12	0
2013/02	52	22
2013/03	4	8

Figure 23. Data view.

qualify them based on other data aspects, such as sentiment value or message media type. In addition, stripping down the dataset obviously removes everything but the most basic elements, so any granular detail that could be gained, like the content of the message, or viewing the actual photo that had been sent, would not be possible. Secondly, this visualization does not show data distributed over a split axis, but rather shows it overlaid on top of the other. While this is acceptable

in cases where there are large differences between the area graphs, in this case the data is pretty equal,

and it would be more beneficial to view it as a split visualization. Finally, the most important realization that came out of visualizing prototype 2 in code was that, as far as I was able to comprehend from my research, it would not be able to create an area graph composed of small units using D3.js. The way that D3.js creates area graphs is as a summation of the data, and not representative of individual data points. This discovery would require a reconceptualization of the prototype plan, and would impact my progression forward.

Reflection on prototype 2

The process of creating this prototype was significantly different than that of prototype 1, which was, in the words of my advisor Ali Qadeer, “code first.” Prototype 2, however, was developed with the intention of answering my research questions, which are grounded in illustrating and understanding the digital dialogue I maintain with my sister on the Facebook messenger platform. Using the research questions as a guide helped me to understand the priorities of the research, and what aspects of the dataset were important to show in the visualization. Applying the visualization design to code was important, as it very crucially allowed me to understand its technical limitations and what would not be possible. At the same time, despite the limitations, I found that even being able to display some aspects of Prototype 2’s code using the dataset from Facebook was meaningful, at least to me; seeing when our conversations on Facebook were frequent, and when they tapered off, is very interesting. I notice patterns in this visualization that I did not necessarily get a sense of from the visualizations created in Tableau, and something that strikes me is noticing the points in time where Zoya’s green area graph exceeds my yellow graph in height; most noticeably right before 2016, and in the middle of 2018. This surprises me, because I was always under the assumption that I was the more talkative sister, and would have expected my area graph to constantly reflect more messages sent than Zoya’s graph. I recall Katherine Davies’ research study “Siblings, Stories and the Self,” which discusses the construction of narratives amongst siblings, and the assignment of contrasting characteristics to siblings in relation to each other. When thinking about my own family dynamics and the way that even my parents describe my siblings and I, there has always been a reinforcement of the notion that Zoya is quieter, and more introverted, while I am the outgoing personality. While there are many points in time where my graph does indeed reflect more messages sent, the numbers are still quite equal. This shows me that we are very responsive to

one another. It seems that in terms of our activity on Facebook messenger, we are very chatty around summertime, and least talkative at the beginning of the year. This makes sense, because during the holidays we're always together, and summer often involves more free time for Facebook chatting. And finally, something I noticed is that the second largest spike in this visualization is from August 2017, the month that I moved away from home. During this month, I sent 874 messages, and Zoya sent 894. The only months that exceed this amount of messages sent are in the following year, August 2018 and September 2018.

Prototype 3

Building on the insights I extracted from the process of planning and developing Prototype 2, I ultimately decided to revise the design, keeping the aspects from the previous prototype that were successful and discarding those that were not. This meant having to move away from the area graph. Although it was still useful in highlighting insights, it wouldn't allow the user to be able to go deeper into the data, and this was something that I was mindful of. Being able to view individual messages would highlight Judith Donath's critique of the hyperpublic within digital spaces (283), as well as helping to enhance the answer to my first research question on how visualization can demonstrate social media's use as a tool to maintain my relationship with Zoya over time and distance. Our connection is the core of the thesis, and only looking at it from an abstracted and generalized perspective is detrimental to a comprehensive understanding of what it encompasses. In my design plan for Prototype 3, I maintained the same axes, with the exception of switching the x- and y-axis from what they represented in Prototype 2. This decision was based on a conversation with my advisor, Isabel Meirelles, who suggested that splitting an axis horizontally often indicates positive and negative values and might be misleading in the case of a comparison of similar values. Thus, in this design, the x-axis is split down the middle, with the y-axis representing time. In addition, Prototype 2 incorporated notation of the distance between Zoya and I as written text placed next to the year it describes. However, Isabel also suggested showing this visually, rather than in writing. I also received assistance from Greice Mariano (a postdoctoral researcher at OCAD and one of my colleagues in the Visual Analytics Lab) in running a sentiment analysis in the programming language R on the message data and assigning sentiment types to each message. This was achieved using the Tidy

sentiment library from R (Silge and Robinson). This really helped in gaining another qualifying aspect of the data that could be used in the visualization, which I planned to employ by use of colour.

Finally, it was necessary to figure out how to show the same essence of information without the area graph. After discussions with both Ali and Isabel, I determined that a good solution might be a mirror histogram, stacked in units to represent each message sent.

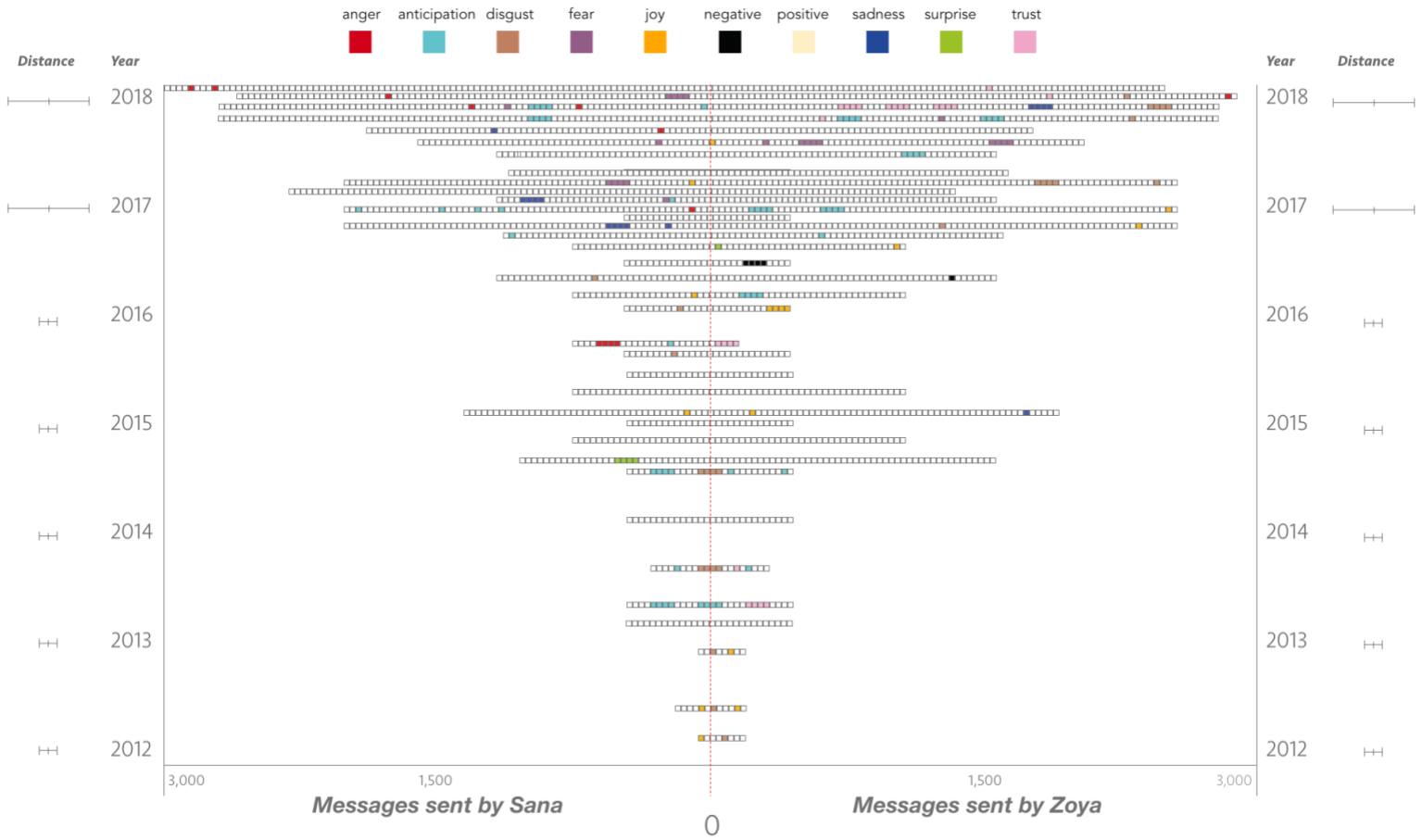


Figure 24. Wireframe of proposed visualization for Prototype 3. Y-axis indicates time, with most recent time at the top. The two x-axes represent numbers of messages sent, and are mirrored according to sender from the center 0 point. Colour indicates sentiment type. Distance is shown iconographically in parallel to the y-axis.

Data sketches

It was important for me to have a chance to experiment with the actual dataset, using at least some aspects of the visualization design with which I planned to approach representing the data in this phase.

Going back to the Tableau software was the quickest way for me to create some of these sketches.

Although it was not ideal to go back to Tableau, attempting to enact this plan in D3.js was taking too

much time. While I still planned to use D3.js in my exhibited prototype, for the purposes of this research I wanted to be able to connect the visualization with the dataset before submitting my thesis document for review. In Tableau, I decided to maintain time as the y-axis and number of messages as the x-axis. I was not able to figure out how to represent each individual message as a unit in this software, but since I was using the tool to get a general sense of the data rather than as the final implementation of the data visualization, I was not too troubled by this. I also was unable to achieve a split x-axis in order to get the mirrored histogram. However, going back to Tableau did allow me to visualize the sentiment type. My first data sketch visualized number of messages, separated by sender, over time, and colour-coded according to sentiment type.

Sentiment categorization was made possible through the use of an R sentiment analysis, which compared all of the words within the content of the message to the Tidy sentiment library—a huge database of words that are categorized into various emotional groups. Messages might be assigned more than one emotional category, but are placed into the group of the emotion that is most prevalent in that message. Neutral messages are those which lack emotionally-charged words. Messages might be angry if the words they contain include expletives. Disgust is a grouping that applies to messages with words like "ew," "gross," "weird," etc. Positive messages include those with words like "like," "fun," "happy," and so forth. Essentially, what becomes problematic with sentiment analysis run by a script that compares message contents to a database of generic emotional words is that Zoya and I have our own specific usage of words and sentence structures. Therefore, the sentiment analysis often showed incorrectly assigned sentiment groupings. This was an issue that had to be fixed in order to accurately assess our messages over time—my solution was to go through the entire dataset by hand to check that all of the messages were put into an accurate grouping.

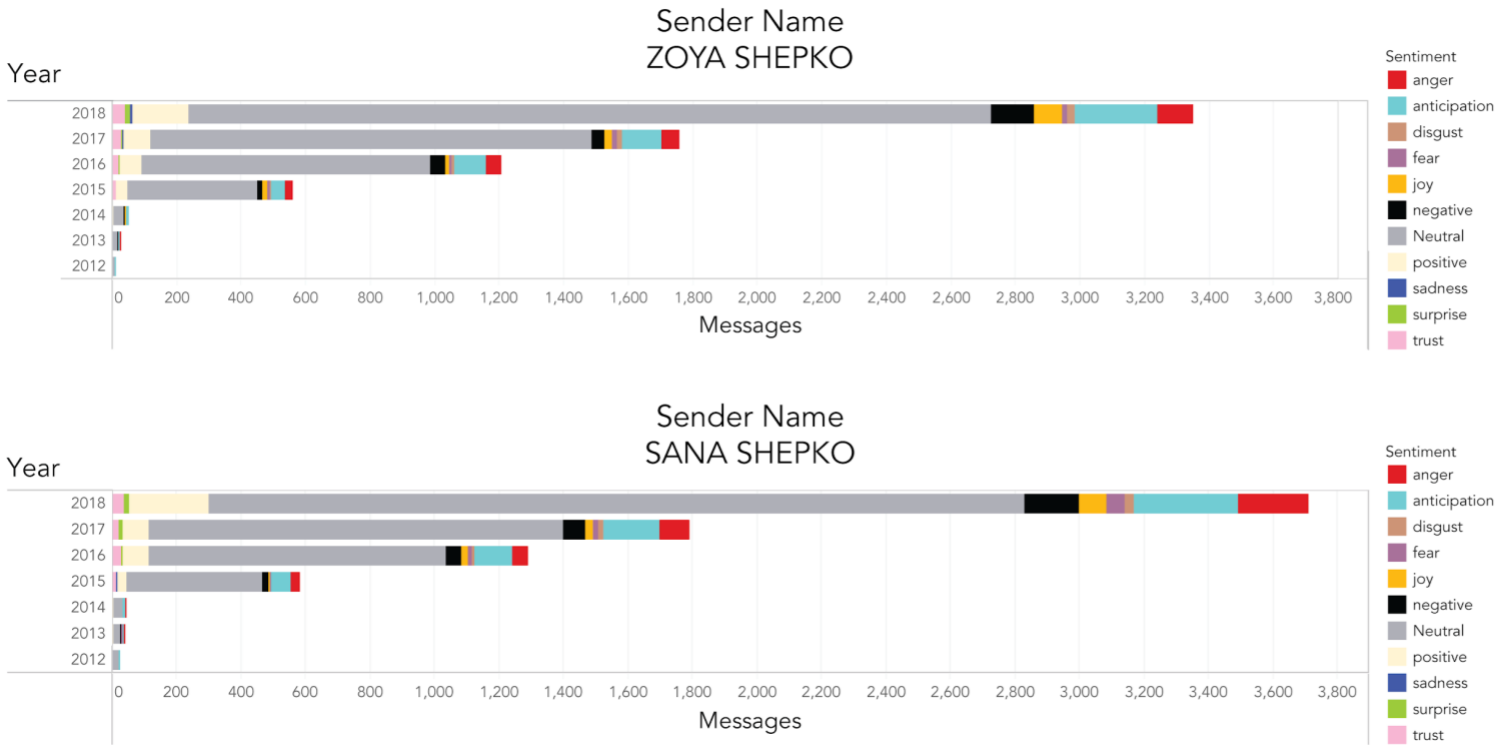
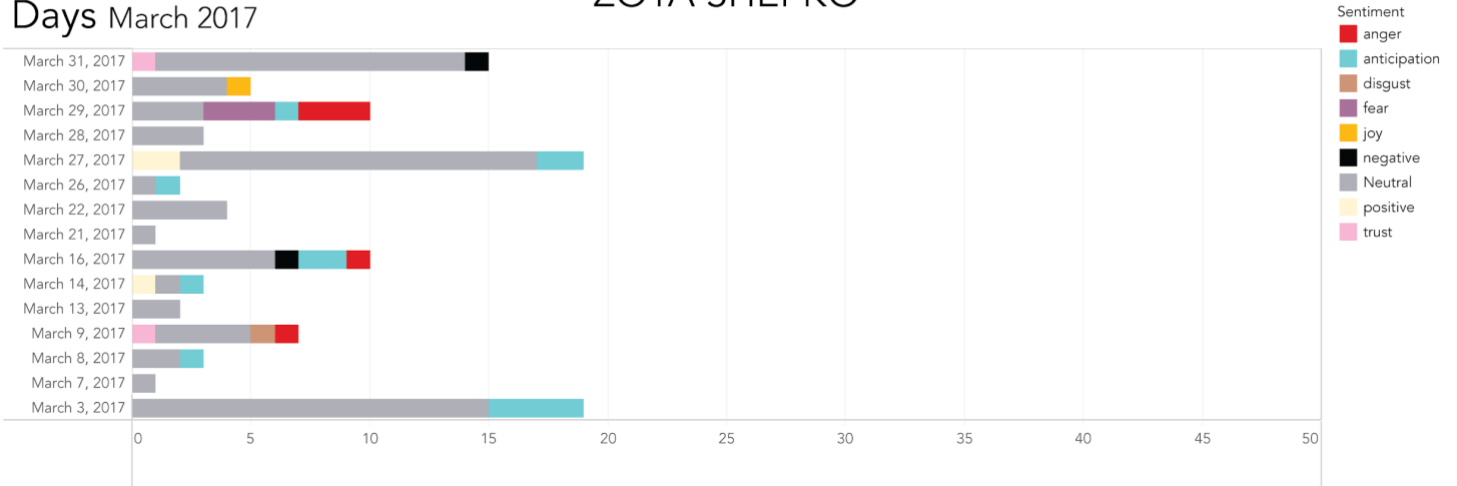


Figure 25. Tableau visualizations showing number of messages sent in each year, graphs separated by sender. Sentiment type is shown through colour.

While I have been making a lot of visualizations of the dataset at a yearly level, I also wanted to see what the visualization might look like on a more granular timescale. Since my research investigates how social media might be used as a tool to maintain a relationship over changing distance and time, I decided to choose a month from before I left home, and compare it to the same month a year later, after moving. I chose the month of March, and compared each sender's March 2017 to March 2018.

Sender Name ZOYA SHEPKO

Days March 2017



Days March 2018

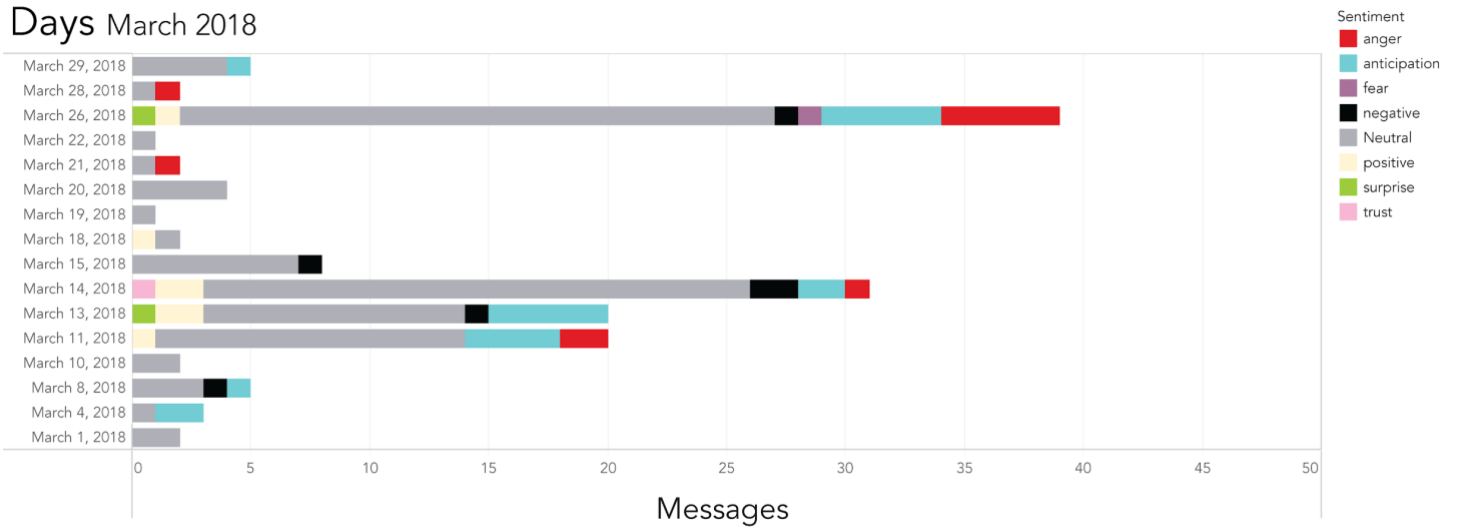
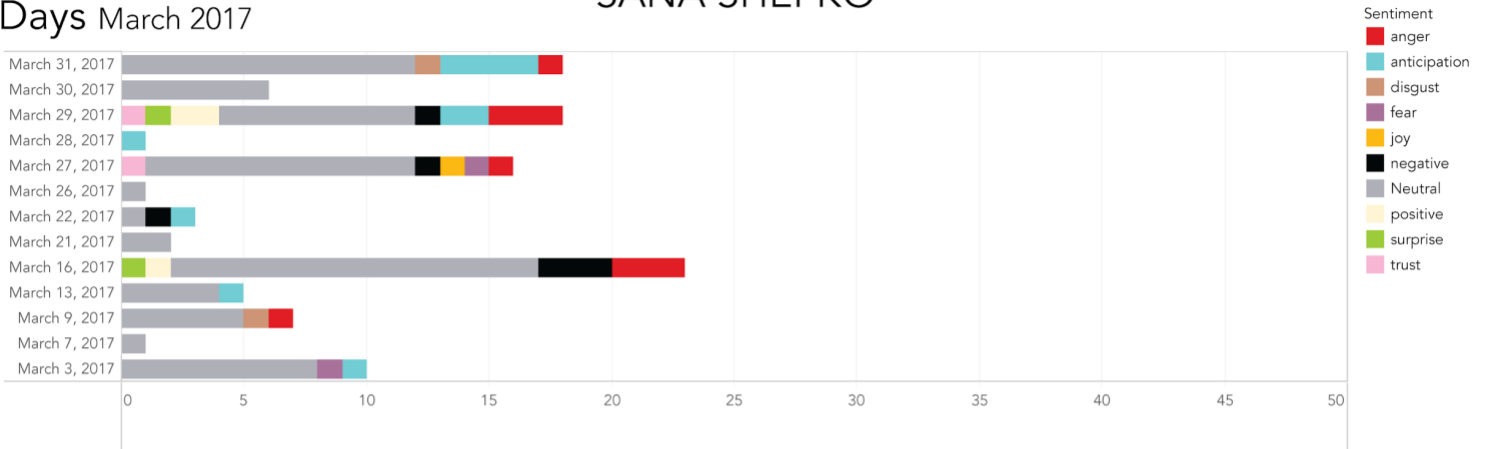


Figure 26. Comparison of graphs showing the number of messages sent by Zoya in March 2017 and March 2018, respectively. Colour indicates sentiment value.

Sender Name SANA SHEPKO

Days March 2017



Days March 2018

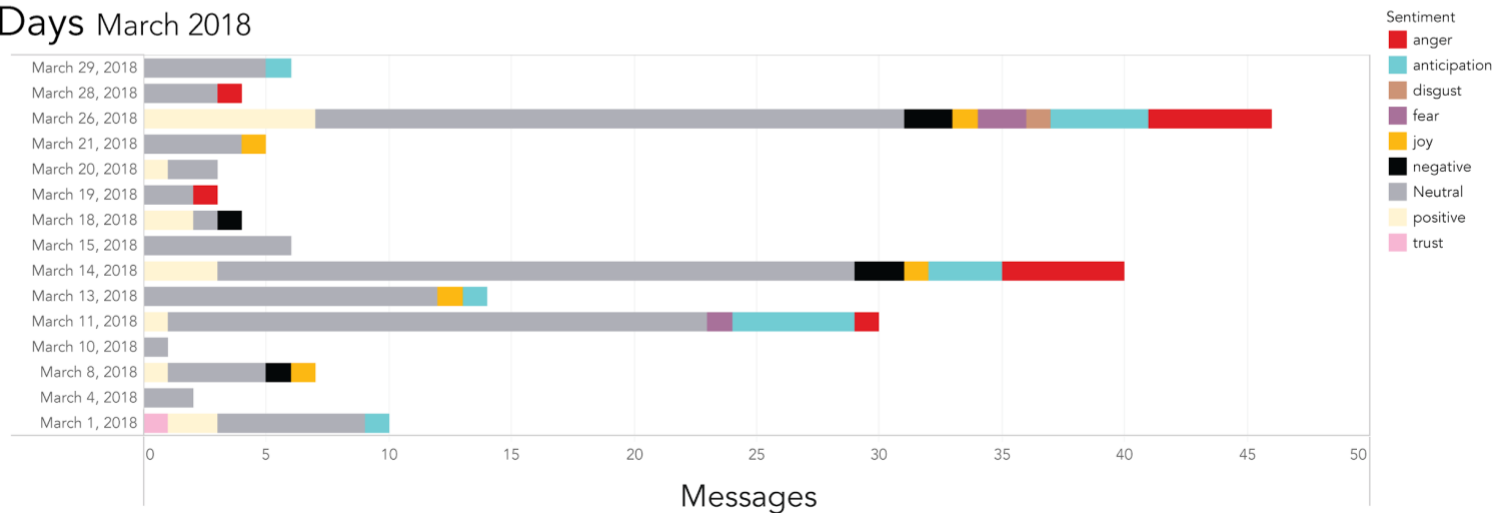


Figure 27. Comparison of graphs showing the number of messages sent by Sana in March 2017 and March 2018, respectively. Colour indicates sentiment value.

The colour used in these visualizations was somewhat influenced by research conducted by D’Andrade and Egan on colours and emotions; their findings are that typically brighter, saturated colours are associated with positive emotions and desaturated colours with negative emotions. I took this into consideration to an extent, while also considering standards of colour usage in

marketing and advertising. It was important that the colours were different enough that it might be easier to distinguish between sentiment values.

It was also useful to see that even before moving, by 2017 Zoya and I were fairly consistent in messaging pretty regularly. It seems that there was never more than a week in either March 2017 or March 2018 without messaging, although in 2018 we were having longer conversations at a time. Extracting from the data sketches in Tableau, I was able to summarize my and Zoya's daily average of messages sent per day in each year. In essence, we went from less than 1 message per day from 2010-2014, to each sending an average of 2 messages a day in 2015. In 2016, I sent an average of 4 messages per day, while Zoya sent an average of 3 messages per day; in 2017, we both sent an average of 5 messages per day; and in 2018, I sent an average of 10 messages per day and Zoya sent around 9 a day.

Coding Prototype 3

Moving from the design into the literal coding of Prototype 3 benefited from having a clear plan of what the eventual visualization should look like. Turning that into an interactive, web-based visualization was extremely challenging (for the sake of brevity I will omit a detailed account of the process of coding), but it was important that I had gained some insights into what was and wasn't possible in D3.js in earlier prototypes as well as a better sense of the ways that visualizing the data might be more meaningful. Therefore, the coded final prototype actually looked similar to the way that I was envisioning it in the design phase. The one exception was that I didn't anticipate the sheer amount of messages that could be sent in a month's time. The wireframe (Figure 24) shows each message as a square. However, because the number sent in a month could be as many as 650 messages, the limited space of the x-axis needed to be taken into account in order to fit the whole visualization within the width of a computer screen. Thus, the coded visualization shows slim rectangles stacked on top of each other in a month, rather than squares. In addition, while Figure 24 shows the geographical distance between my sister and I in the form of icons, I was unable to implement that in the coded version due to the lack of space.

Some additional components that I had not necessarily thought about in the design phase were related to the user experience and interactivity of the site. In the coded version, the visualization is actually quite long, and in order to reach the beginning of the message data from 2010 the user has to scroll very far

down the screen. A user who hadn't noticed the titling of the left and right side of the mirror graph might be confused which side belonged to which sisters' messages once they had scrolled past the titles and legend and were in the midst of the visualization. Because of this realization, I decided to make the titles "messages from zoya" and "messages from sana" stick to the top of the screen as the user scrolls down the visualization.

Another UX component that evolved out of coding this prototype was the idea of filtering out messages in the visualization by interacting with the sentiment group legend. The way I was able to achieve sentiment filtering was through a toggle-like interaction; i.e., when the website first loads, all of the sentiment groups are automatically toggled 'on', and if a user were to click on one of the sentiments, like "neutral," it would toggle off, and remove all of the messages categorized as neutral from the visualization. Because of how compressed the visualization is in width, filtering makes it easier to isolate messages within the column. It also makes it possible to view and interact with the dispersion of emotional types in our messages over time. For example, toggling off all of the sentiments except for "sadness" would show the points in time when Zoya and I were sending messages that were categorized as being emotionally sad.

Finally, the interactivity of hovering over a message to read its contents was something that I had planned from the very beginning. Fortunately, I was able to figure out how to make this happen in this visualization, as well as display the date and time of day in which that message had been sent. Another aspect that I decided to include was to match the background colour of the tooltip to the sentiment type of the message it was showing. This would identify the sentiment type of the message being viewed through the tooltip.

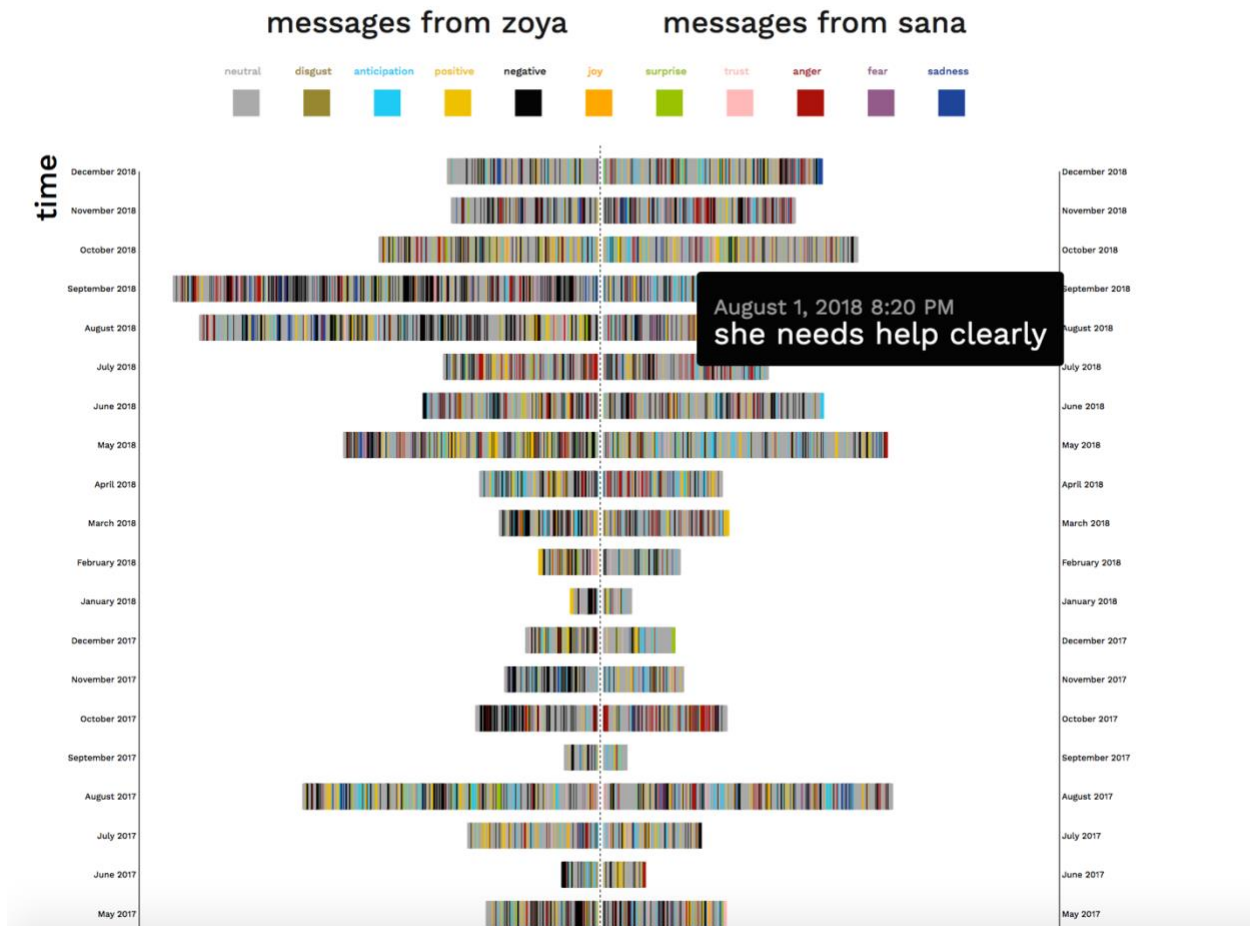


Figure 47. Screenshot of the coded visualization from Prototype 3.

I would like to note that, when working on the datasets to create the coded visualization, I realized the R script sentiment analysis that had been completed earlier was not as accurate as I had hoped. Therefore, it was necessary for me to go through the datasets by hand in order to ensure that each message was assigned a correct sentiment group. Going through every message by hand also served as a way for me to reflect on the entirety of our message history on Facebook and relive the progression of our conversations through time. This was helpful to me as a general reminder of the events occurring in our lives. Visually, the colouring of the messages make it possible to identify the moments in time that were dense in a certain emotion type, and hovering over the messages provides a little more context of what these messages were regarding; however, having read through the entire dataset and refreshing my memory of our conversational history, I am able to understand more holistically the circumstances that led to clusters of emotional colours.

Thesis Installation

For the Digital Futures graduate thesis exhibition, I ran the data visualization website off a local server on my computer and connected it to a large monitor. I felt it would be important to be able to see the visualization all at once rather than to be limited to scrolling through it on the screen, so I created two large posters, split into messages from Zoya and messages from me, that showed the entire span of messages over time. These posters included guiding annotations above each axes and above the sentiment legend to help lead a viewer through understanding how to read the chart. Each time frame in the time axes had an annotation, summarizing the big events of that time and what the gist of the message data at that time was about. This allowed the project to be more contextualized in the lived experiences of Zoya and I rather than an abstract visualization.

December 2018

Zoya and my partner plan a surprise birthday party. I go home for a couple weeks.

November 2018

I am stressed out this month, and am sad about being away from home for my birthday.

October 2018

Zoya and our brother come to visit my partner and I for Canadian Thanksgiving. I go to Berlin for a week.

September 2018

Zoya is stressed out this month, and is having a hard time. She gets frustrated with me. I ask her to be part of my thesis.

August 2018

Zoya works at a music camp. I consider coming to visit her, but back out. Zoya gets a full-time job.

July 2018

We have multiple performances this month. I am home for a little while. I feel sad and uncertain.

June 2018

Our friend gets married this month. We discuss logistics. I go to Paris with my partner.

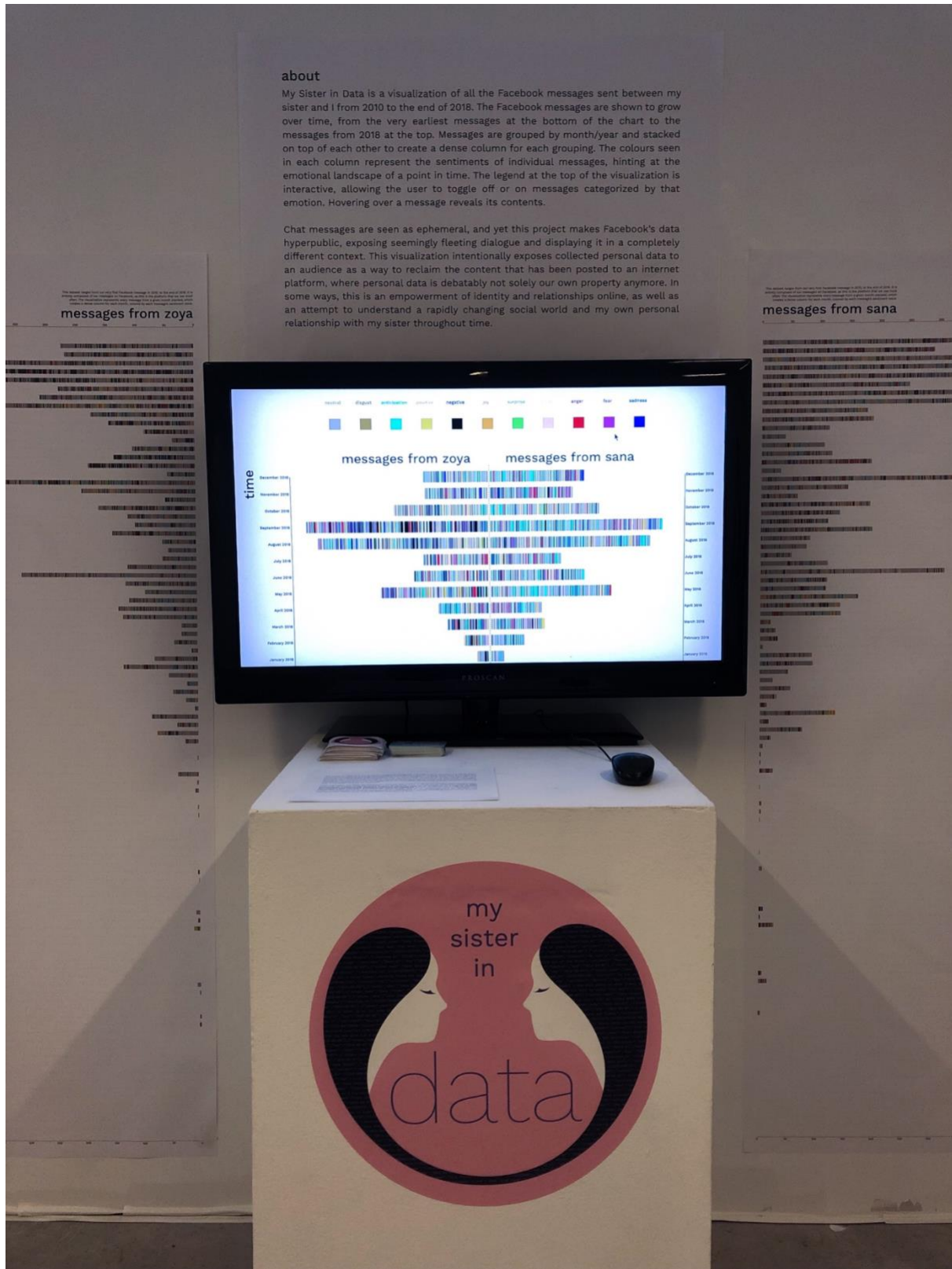
May 2018

Zoya graduates this month, I come home. She comes to Canada for my partner's birthday and to visit me.

April 2018

I finish my second semester of graduate school. We discuss upcoming weddings.

Figure 48. Close-up of the time axis annotation on the posters in my thesis exhibition.



about

My Sister in Data is a visualization of all the Facebook messages sent between my sister and I from 2010 to the end of 2018. The Facebook messages are shown to grow over time, from the very earliest messages at the bottom of the chart to the messages from 2018 at the top. Messages are grouped by month/year and stacked on top of each other to create a dense column for each grouping. The colours seen in each column represent the sentiments of individual messages, hinting at the emotional landscape of a point in time. The legend at the top of the visualization is interactive, allowing the user to toggle off or on messages categorized by that emotion. Hovering over a message reveals its contents.

Chat messages are seen as ephemeral, and yet this project makes Facebook's data hyperpublic, exposing seemingly fleeting dialogue and displaying it in a completely different context. This visualization intentionally exposes collected personal data to an audience as a way to reclaim the content that has been posted to an internet platform, where personal data is debatably not solely our own property anymore. In some ways, this is an empowerment of identity and relationships online, as well as an attempt to understand a rapidly changing social world and my own personal relationship with my sister throughout time.

The actual legend from the My Sister in Data Facebook message visualization is as follows: neutral, disgust, satisfaction, positive, negative, no, surprise, love, anger, fear, sadness

messages from zoya

The actual legend from the My Sister in Data Facebook message visualization is as follows: neutral, disgust, satisfaction, positive, negative, no, surprise, love, anger, fear, sadness

messages from sana

Figure 49. A picture of the full thesis installation setup.

Reflection on Prototype 3

Overall, the data visualization very clearly shows how over time, messages between Zoya and I steadily increase. In the earlier years of the dataset, from around 2010-2015, we would go months at a time where there were no messages at all sent. From 2015 onward, we use Facebook to chat regularly, and there are no longer gaps in the visualization. Adding the additional data of the sentiment value was also interesting and helped me learn that while neutral messages are always the most common, we also display a wide variety of other emotions. Finally, the data visualization shows that we are very responsive to one another. It is a very mirrored visualization, which means that we are having a steady dialogue.

The sociological articles by Thomas Lee and Katherine Davies were most directly applicable to my autoethnographic methods of self-reflection and situating my own experiences as part of the research. Using their theories and research and applying it to my relationship and communication with my sister made me notice things I would not have before; for example, Thomas Lee's study on communication levels between siblings notes that communication is frequent in the case of emotional closeness or geographical proximity. In the year after I move away, the communication between my sister and I actually increases. Interpreting this through Lee's theory suggests that this may be due to emotional closeness. Earlier in the process of creating data visualizations, I noted that I was a bit surprised when I first noticed that visualizing the data showed that Zoya was as chatty if not more so than I was during certain time frames, as I had always considered myself the more talkative sister and her the quieter one. This phenomenon is noted by Katherine Davies' research on construction of narratives about sibling identities in families.

Finally, I found that the installation of this prototype and watching people interact with the interface led to new insights. Many people mentioned that they felt voyeuristic reading through the messages, and as if they were violating my privacy. This was interesting to me, as I think I have desensitized myself at this point so much that allowing private messages between my sister and I to be accessed by strangers doesn't phase me anymore.

conclusion

This thesis is an exploration of relationships on the internet, focusing on the interaction between my sister and I on Facebook Messenger from 2010-2018. My thesis research is categorized into three different categories: the process of my first, second and third prototype designs, each offering their own unique insights to this research.

Through the research of Prototype 1, it has been possible to identify trends in the language used between me and my sister, and common phrases. In addition, it is possible to note the sorts of situations which inspired common messages. For example, Zoya sent 'Omg' as the exact content of a message to me approximately 99 times. This is usually directly in response to messages that I have sent to her, rather than an instance of 'double-texting' (to use colloquial terms). These messages from me tend to either be melodramatic (sometimes in all-caps), or when I am teasing her: 'WHY CAN'T [name] DROP HER OFF,' and 'Are you finding wayfe zoy' ("wayfe" in this case is also a phrase that has become common between the two of us and our shared close friend group, and indicates a romantic partner, so in this case I'm teasing her).

In 2010-2013, the earliest years of the dataset, the interaction is at its lowest; while in 2018, the messages sent back and forth are at their highest amount. It is unclear if this is necessarily due to moving away from each other in 2017, although research from Prototype 2 shows that August 2017, the month I move to Toronto, contains the third highest number of messages sent in any month overall, with only the following August 2018 and September 2018 surpassing that amount of messages in a month.

The process of Prototype 3 introduced a new layer of qualifying data to the visualization, a sentiment analysis that was made possible with the assistance of OCAD postdoctoral researcher Greice Mariano. The messages are shown as a mirrored histogram, visualizing the symmetry of online dialogue through time. Each message is stacked by month, and coloured by sentiment. The sentiment legend is interactive, allowing the user to toggle off or on messages based on emotion, and hovering over individual messages reveals the contents of the message and the date and time it was sent. The decision to assign these colors to certain sentiments were guided by the research article "The Colors of Emotion" by D'Andrade. This research found that darker, low saturated colors were associated with negative emotions

while bright and high saturated colors with positive emotions. The implementation of sentiments for each message was important, as it allowed me to create a toggle user interaction and add a more intimate understanding of the data, while the user interaction of hovering over the visualization in order to read message contents highlighted Judith Donath's notion of hyperpublicity. I found that before 2015 Zoya and I both send less than one message a day, but after 2015 messages steadily increase. In 2017, we both go from 5 messages a day on average to double that in 2018, with me sending around 10 messages a day and Zoya sending around 9 messages a day.

Implications

The research shows, most importantly, that there is notable increase in interaction over time. I believe that this helps to partially answer my first research question, as this suggests that social media is indeed used to maintain the relationship. Part of this research question is how this is impacted by time and distance; an important aspect to my research through design process is identifying the differences in the visualizations of the time before and after I move away from home in 2017. In the research of Prototype 3, I find that the average message per day in 2017 is at 5 messages for both Zoya and I. The average number of messages I send in 2018 doubles from the year prior, at 10 messages per day, while Zoya similarly increases at an average of 9 messages per day in 2018.

However, it is also crucial to highlight that most people's social media usage has grown over time. In 2018 more than 3 billion people around the world used social media at least once a month, while the annual growth global average of social media users is 13% (Kemp). The growth of people's usage of social media is reinforced in Sherry Turkle's "Reclaiming Conversation," where she discusses the phenomena of technology taking away from people's quality of time and conversation together. Students, going out to dinner with their friends, will have their phones on the table and don't consider it rude to be on their phones as long as there are at least 3 people carrying on the conversation. In this example, Turkle writes that the most commonly heard phrase at these get-togethers is, "wait, what?" when someone picks up their head from their phone and tries to jump into the in-person conversation without having followed the previous narrative. As a young person having gone out with people my age, I can attest that this is definitely something that happens. I myself am a frequent utterer of "wait, what?" in many conversations, and it is indubitably because I had been distracted by some other thread of

conversation or interaction on my phone. These reflections on social media usage, combined with the results from this thesis, demonstrate the expansion of digital tools and interfaces as mediators of social relationships, and the development of the posthuman world. Considering that one of this thesis' research questions asks how changing distance impacts the use of social media to maintain our relationship, while global statistics show that social media usage steadily grows over time, it is difficult to accurately distinguish how distance specifically impacts the frequency of our messages. Therefore, within the scope of this research it was not possible to uncover this aspect of the research question.

I believe that it is important to emphasize that this thesis intends to be an example of an existing and ever-growing cultural phenomenon. The fact that we have been using a social media platform to such an extent in order to communicate, rather than solely relying on more obvious forms of communication such as texting or calling, indicates how much of a presence social media has in our lives. It seems to me that I would often use Facebook to message Zoya because it was more convenient. I often have Facebook open on my laptop, and moving from scrolling through the website's "Timeline" to simply opening up the chat with her on Facebook seemed much more efficient than having to get my phone to text her. I also believe that a lot of our interaction was contextualized within Facebook; we often went off into our private chat to discuss ongoing conversations within other Facebook group chats we were both participants of, or to send screenshots of posts from other people on Facebook.

Research from Daniel Miller et. al notes that users of social media seek out "niches" of social relationships within various platforms, and perform according to the audience to whom they are presenting. Considering how Zoya and I's Facebook messages often refer to the other social groups we locate ourselves within across other areas of Facebook, I find that Miller's idea of social niches resonates with me. I believe that in the case of a very close familial relationship enclosed within the space of our private messages on Facebook, the idea of performativity becomes less obvious than if I had used the data from all the posts I had made for all my Facebook friends over time. However, I keep coming back to my fascination with the documentation and archiving of online interactions, and how the impermanent personality and relationship that existed when I was 16 and Zoya was 14 has transformed over time. When I sent the visualization and datasets for Zoya to review, the first thing she messaged back to me was: "Woosooooowwww you were meannnnn to me I just wanted to be your friend lol." Her remarking on

the relationship we used to have when we were younger highlights how different our interactions are now. Personal data becomes so interesting to me because it makes permanent the aspects of identities that are otherwise completely amorphous. Additionally, I feel that the performance of identity online is very structured and limited by the affordances of the interface itself; Facebook as a platform has grown in the functionality it allows its users, but in the early 2010s messaging through Facebook was most commonly used to send text or emojis. The digital tools at our disposal to express ourselves are restricted to their contemporary technological state. Therefore, the particular case of a conversation over time between two sisters on Facebook documents the performances of our metamorphic personalities as we grow from teenagers to young adults, as well as the changing affordances and constraints of using a social networking platform as our main form of non-physical communication. This has been made clear to me through the process of going through the data by hand, and it must be stated that the visualizations themselves would not immediately unveil such insights; rather, it is because of my own deep-dive into the dataset that I am able to uncover these discoveries.

Ultimately, the process of researching this topic has been insightful, heartwarming, and also bittersweet for me. My homesickness is sometimes alleviated, and sometimes heightened as I work through the message dataset. I worry that my own experience is not always easily translatable to an audience outside of my perspective. However, this is something that I have grown to be aware of, and hopefully I will have opportunities to improve the communication of my perspective in the future.

Future work

During the graduate thesis exhibition, several people who had interacted with the data visualization mentioned to me the potentiality of using this to understand other relationships, as a therapeutic and reflective tool. For example, one suggested visualizing conversations and categorizing by emotion could be used in the context of couples therapy, with the potential of having a more general view of how the conversation changes over time. Using data visualization as a form of therapy is definitely intriguing, and reinforces the importance of relationships between people. This might not need to be necessarily using data from an online platform, but could also include transcripts of actual conversation.

From a more UI/UX standpoint, I believe that there are many more directions that this visualization could take that would allow a user to dig deeper into the huge amount of data available. For

example, in the design of Prototype 2 I initially wanted to include the ability to change timescales. At its current state in Prototype 3, the visualization is fixed on a monthly timescale. However, if a user were able to adjust the y-axis settings, they would be able to see the visualization across a yearly, weekly, or daily level. This functionality would make it possible to have more control over the visualization and to isolate specific points in time.

Another aspect that I would like to include is related to my inspiration from Jonathan Harris' *We Feel Fine* project (2005). Harris' work combs through blog entries online using the words "I feel" or "I am feeling" to gather data on human emotion across the internet. The visualization incorporates various levels of interactivity, and there are many ways for a viewer of the visualization to drill into more details and statistics. Drawing from this project, another aspect I would like in this visualization is the ability to click onto the message itself and be taken to a different screen that displays additional statistics and information, perhaps bringing back elements from Prototype 1 related to language and repeated message contents. For example, clicking on a message could show the user how often we use each word that is in that message, or how that message is connected to a specific dialogue.

bibliography

- Anderson, Laurel and Deborah Brown McCabe. "A Coconstructed World: Adolescent Self-Socialization on the Internet." *Journal of Public Policy & Marketing*, vol. 31, no. 2, 2012, pp. 240-253.
- boyd, danah m. and Nicole B. Ellison. "Social Network Sites: Definition, History and Scholarship." *Journal of Computer-Mediated Communication*, vol. 13, no. 1, 2007, pp. 210-230.
- Brigg, Morgan, and Roland Bleiker. "Autoethnographic International Relations: Exploring the Self as a Source of Knowledge." *Review of International Studies*, vol. 36, no. 3, 2010, pp. 779-798.
- Cadwalladr, Carole and Emma Graham-Harrison. "Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach." *The Guardian*, 2018.
- D'Andrade, R., and M. Egan. "The Colors of Emotion." *American Ethnologist*, vol. 1, no. 1, 1974, pp. 49-63. *JSTOR*, www.jstor.org/stable/643802.
- "Data Policy." *Instagram Help Center*, Instagram Inc., 2018.
- Davies, Katherine. "Siblings, Stories and the Self: The Sociological Significance of Young People's Sibling Relationships." *Sociology*, vol. 49, no. 4, 2015, pp. 679-695., www.jstor.org/stable/44016699.
- Davis, Jenny L. "Triangulating the Self: Identity Processes in a Connected Era." *Symbolic Interaction*, vol. 37, no. 4, 2014, pp. 500-523.
- Donath, Judith. *The Social Machine: Designs for Living Online*. The MIT Press, Cambridge, MA. 2014.
- Dreyfuss, Emily. "Teens Don't Use Facebook But They Can't Escape It, Either." *Wired*, February 2019.
- Frenkel, Sheera et. al. "Delay, Deny and Deflect: How Facebook's Leaders Fought Through Crisis." *The New York Times*, 2018.
- Goffman, Erving. *The Presentation of Self in Everyday Life*. Doubleday, New York, 1959.
- Greenwald, Glenn, et. al. "Edward Snowden: the whistleblower behind the NSA surveillance revelations." *The Guardian*, 2013.
- Gruzd, Anatoliy, et. al. "The State of Social Media in Canada 2017." *Ryerson University Social Media Lab*, 2018.
- Harris, Jonathan, and Sep Kamvar. *We Feel Fine*. 2005.
- Hayles, N. K. *How We Think: Digital Media and Contemporary Technogenesis*. University of Chicago Press, Chicago, 2012.
- Huyke, Hector J. "Toward an Ethics of Technologies as Prostheses." *International Journal of Technology and Design Education*, vol. 11, no. 1, 2001, pp. 53-65.
- Kemp, Simon. "Digital in 2018: World's Internet Users Pass the 4 Billion Mark." *We Are Social*, 2018.
- Lee, Thomas R., et al. "Sibling Relationships in Adulthood: Contact Patterns and Motivations." *Journal of Marriage and Family*, vol. 52, no. 2, 1990, pp. 431-440. *JSTOR*, www.jstor.org/stable/353037.

- Lovink, Geert. "After the Social Media Hype: Dealing with Information Overload." *e-flux*, vol. 45, 2013.
- Manning, Erin. *Politics of Touch: Sense, Movement, Sovereignty*. University of Minnesota Press, 2006.
- Manovich, Lev. *The Language of New Media*. MIT Press, Cambridge, MA, 2001.
- Medrek, Ania. *News by Association: Designing a way out of the echo chamber*. 2018. OCAD University, MDes.
- McLuhan, Marshall. *Understanding Media: The Extensions of Man*. McGraw-Hill, 1964.
- Miller, Daniel, et. al. *How the World Changed Social Media*. UCL Press, 2016.
- Pask, Gordon. "The Background of Cybernetics." *An approach to Cybernetics*. Hutchinson & Co Ltd Publishers, 1961.
- Postman, Neil. *Amusing Ourselves to Death: Public Discourse in the Age of Show Business*. Penguin Books, New York, 1986.
- "Privacy Policy." *Snap Inc.*, 2018.
- Reiss, Paul J. "The Extended Kinship System: Correlates of and Attitudes on Frequency of Interaction." *Marriage and Family Living*, vol. 24, no. 4, 1962, pp. 333–351. *JSTOR*, www.jstor.org/stable/349102.
- "Relationship." *OxfordDictionaries.com*. Oxford Dictionaries, 2018. Web. 21 November 2018.
- Schneier, Bruce. *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World*. W W Norton & Co Inc, New York, 2016.
- Shakya, Holly B. and Nicholas A. Christakos. "A New, More Rigorous Study Confirms: The More You Use Facebook, the Worse You Feel." *Harvard Business Review*, April 10 2017.
- Silge, Julia and David Robinson. *Text Mining with R: A Tidy Approach*. O'Reilly, 2019.
- Smith, Aaron and Monica Anderson. "Social Media Use in 2018." *Pew Research Center*, 2018.
- Srnicek, Nick. *Platform Capitalism*. Wiley, 2016.
- Terranova, Tiziana. "The meaning of Information." *Network Culture: Politics for the Information Age*. Pluto Press, 2004.
- Thurm, Scott. "15 Years of Facebook in 4 Charts." *Wired*, February 2019.
- Turkle, Sherry. *Alone Together: Why We Expect More from Technology and Less from Each Other*.
- Turkle, Sherry. "Computational Technologies and Images of the Self." *Social Research*, vol. 64, no. 3, 1997, pp. 1093-1111.
- Turkle, Sherry. *Reclaiming Conversation: The Power of Talk in a Digital Age*. Penguin Books, 2016. Basic Books, New York, 2011.
- Wellman, Barry and Keith Hampton. "Living Networked On and Offline." *Contemporary Sociology*, vol. 28, no. 6, 1999, pp. 648-654.
- White, Lynn K., and Agnes Riedmann. "Ties among Adult Siblings." *Social Forces*, vol. 71, no. 1, 1992, pp. 85–102. *JSTOR*, www.jstor.org/stable/2579967.

Whiteman, Shawn D., and Abigail Christiansen. "Processes of Sibling Influence in Adolescence: Individual and Family Correlates." *Family Relations*, vol. 57, no. 1, 2008, pp. 24–34. *JSTOR*, www.jstor.org/stable/40005365.

Wiener, Anna. "What It's Like to Wallow in Your Own Facebook Data." *The Atlantic*, Sep. 2018.

Wolfson, Sam. "Snapchat photo filters linked to rise in cosmetic surgery requests." *The Guardian*, Aug. 2018.

Zuckerberg Responses To Judiciary Committee Questions For The Record." *United States Senate Committee on the Judiciary*, 2018.

appendices

The appendix is split into two subsections. Appendix A includes the complete documentation of my initial autoethnographic approach of tracking and screenshotting my weekly Screen Time usage on my iPhone, and using this as a technique to prompt my autoethnographic reflection of the past week.

Following the autoethnographic Screen Time documentation, Appendix B includes the redesigned accessible versions of the data visualizations from the body of the thesis, reproduced to be viewed in black and white.

Appendix A - Screen Time and writing

The following documentation of my iPhone's weekly Screen Time usage was initially used as a technique to discuss my relationship with technology. However, the time-frame of the Screen Time tracking did not correspond for more than a couple of months of data from the Facebook dataset; in addition, it directed my autoethnographic writing to be more a discussion of my personal relationship to technology rather than reflecting on my relationship with my sister. Ultimately, it became less important to my thesis, and therefore was removed from my process chapter and instead relocated to the appendix.

Screen Time - Introduction

Screen Time is a feature released in September 2018 as part of Apple's iOS 12 update. It allows iPhone, iPad, and iPod touch users to receive a weekly report and breakdown of activity and device usage. It also gives Apple device users the option of controlling the amount of time they spend on their devices through the Downtime and App Limits features. Screen Time, as a preliminary step towards an acknowledgement of the immense of time spent on digital devices, was a compelling tool to use as an autoethnographic technique to help to understand my own relationship to screens. However, it is important to highlight that at this point, Screen Time only makes it possible to track the use of mobile devices, and does not include data from the use of other devices like laptops or desktop computers (I shudder to think what my results would be like if it did).

I settled on documenting my personal experiences in writing, on a regular weekly basis. This writing was prompted every Sunday by an alarm that I set on my phone, at 9:30 PM, to screenshot my weekly Screen Time results. Beginning on November 4, 2018, I took screenshots of my Screen Time report for the week leading up to that day. The attached images show the screenshots from the first Sunday of the months of November 2018, December 2018, January 2019, and February 2019.

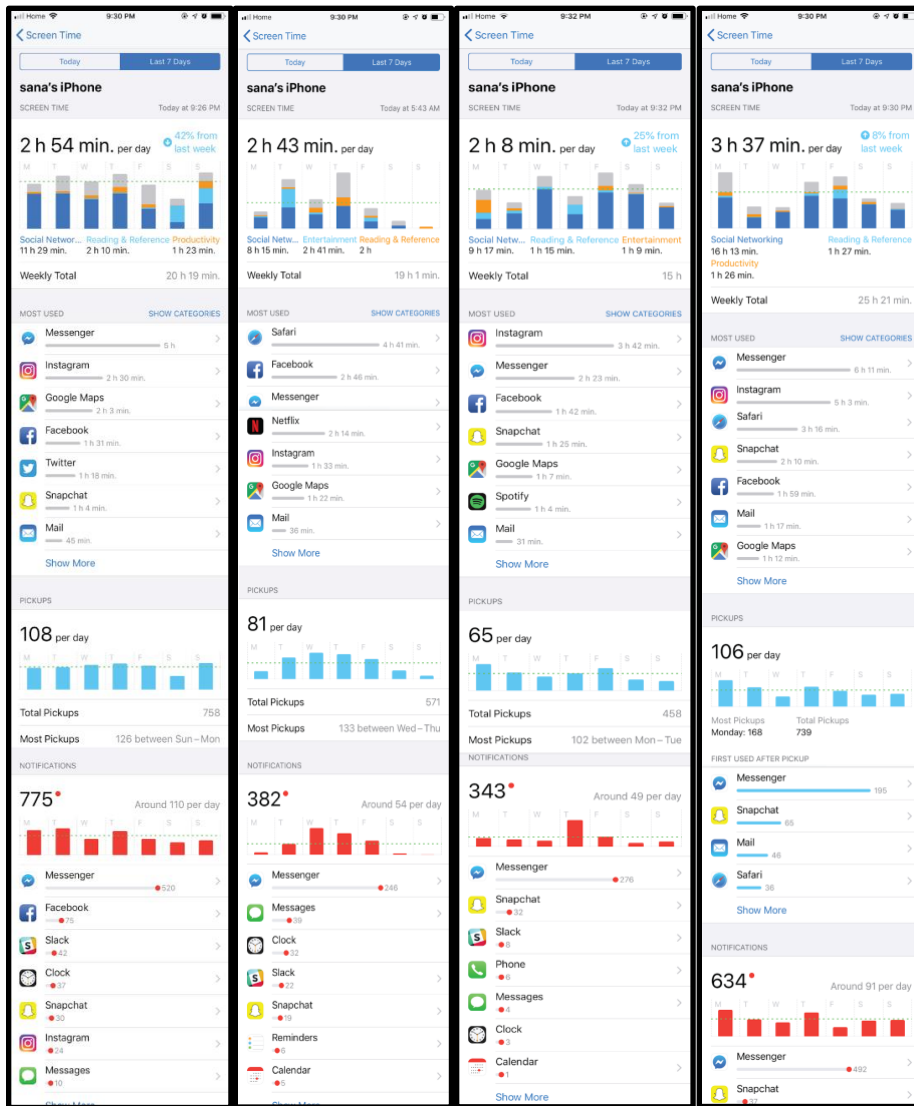


Figure 2. Nov. 4, 2018 Figure 3. Dec. 2, 2018 Figure 4. Jan. 6, 2019 Figure 5. Feb. 3, 2019

Some weeks showed significantly more usage than others, and I was usually able to identify days when I was unoccupied (for example, during the winter break in December there was a huge increase in my use of the Netflix app) as well as the times when I was doing a lot of communicating with others (leading to an increased use of the Messenger app). The process of documenting my phone usage also led to a

heightened awareness of the extreme presence that digital technology had in my life; on December 16, 2018, I wrote:

Also, I notice even while writing this I have this urge to turn on and check my phone every couple of minutes. I know that I haven't received any notifications since my phone would vibrate if I did; and yet, if there's any moment where I require my brain to be working a little bit more focused than normal I put off the moment where I actually have to work by being like 'ok time for a quick phone break and then I'll write about all the complicated stuff.'

Sherry Turkle, in *Reclaiming Conversation*, writes that even the mere presence of a phone sitting within the user's line of sight can lead to distraction, to the point that it affects people's ability to have a conversation together. "If we think we might be interrupted, we keep conversations light, on topics of little controversy or consequence... each feels less connected to the other than when there is no phone present. Even a silent phone disconnects us" (21). The presence of personal devices interrupting in-person interaction is something I noticed when I was home during winter break (written on December 16 as well):

Speaking of hanging out with my sister in person this past week, I also find it humorous that, while I'm thinking about this thesis and me and my sister's relationship online, even when we hang out in person we sometimes fall into this habit of both being on our phones while being physically together. The internet beckons constantly, even when the people we actually want to spend time with are right in front of us. I will also note that being engaged in conversation or just consciously present with another person takes energy. It is so much easier to be on your phone, distracted and being fed content that doesn't really necessitate a lot of mental consciousness, than to think of what to say to another person in front of you. Zoya wakes up really early for work during the week and comes home after dark (the sun sets around 4:24pm this time of year in New Paltz, New York) and I know how tired she must feel. I don't hold it against her that for her to spend time with me sometimes means that it's easier to sit together and be on Instagram than to sit together and just talk. I think I'm also very open in a way that she's not always comfortable with, and a lot of our interaction is sometimes just me talking to her and her telling me whether or not I'm being an idiot. She's a bit more private than I am.

One thing that I find interesting when reflecting on this collection of my Screen Time usage is the reinforcement of how much the Messenger app plays a part in my phone usage. In almost every week's recorded Screen Time, Messenger is the app I've spent the most time on, averaging at about 4 hours of usage per week. It's also the app that I receive the most notifications from, ranging from a minimum of 50 notifications for the week to around 600 notifications received from Messenger for the week. At the end of January 2019, Screen Time added another category of usage tracking that indicates the first app checked after picking up the phone; indubitably, the Messenger app is the first app I almost always check after turning on my phone. This validates my choice in using the data from Facebook Messenger, as this truly appears to be the app that enables most of my social interactions.

Screen Time - Documentation

November 4, 2018

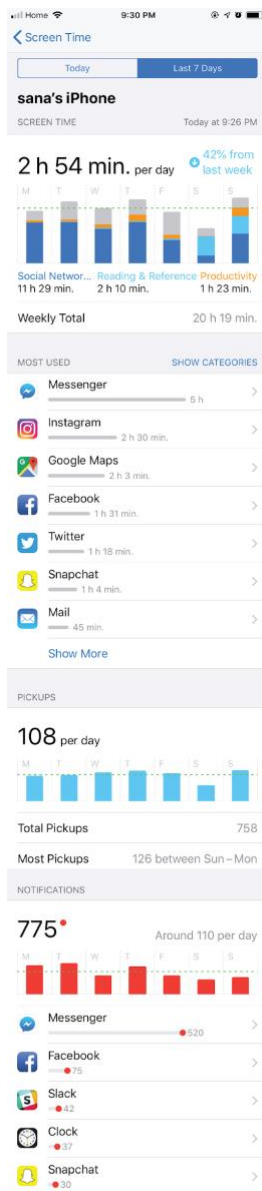


Figure 28. November 4, 2018.

I take this data, or screenshots of pre-existing visualizations of this data, as a way to claim accountability for my relationship with technology and with my devices.

This is a way for me to begin my autoethnographic process, and it seems appropriate for me to continue this method through the use of investigating personal data collected through the entertainments I indulge in on a daily basis—or in another perspective, the technology I rely on almost constantly in order to get through my life. This week, I have an average of around 3 hours per day on my phone. This seems to make sense; I was mostly working and at school and not really socializing with real live people so it only seems appropriate that I would spend 3 hours a day staring at a screen trying to feel connected to others.

Though I need to clarify that the Screen Time function, at least at this point in my research, is only available on iPhone and iPad, so it's not even taking into account all of the time I spend on my computer, which I'm sure would make the amount at least 6 hours a day on a screen. Another thing that makes me interested is this idea of Screen Time as primarily a way to monitor [children's tech usage](#). Why the kids? Aren't enough adults wasting their time online? I certainly am, and feel like I'm constantly aware of the total loss of control that I experience when seduced by the distraction that the digital offers. It really is such an illusion of comfort, and it's undubitably entertaining, but at the same time the underbelly is dark and unnerving. To say that I enjoy living in an age where the mark of social events is people staring at lit screens is a lie. Also, I acknowledge the irony

in all of this. My discrediting the services that allow me to stay connected with family every day; I get it. I'm complaining too much, maybe. But at the same time, I also notice that there is something unhealthy in this, and I don't know that it is always a service or a disservice. Anyways, I hope that this autoethnography turns into something less rambly but I fear at this point in my fatigue it will basically be like this.

November 11, 2018

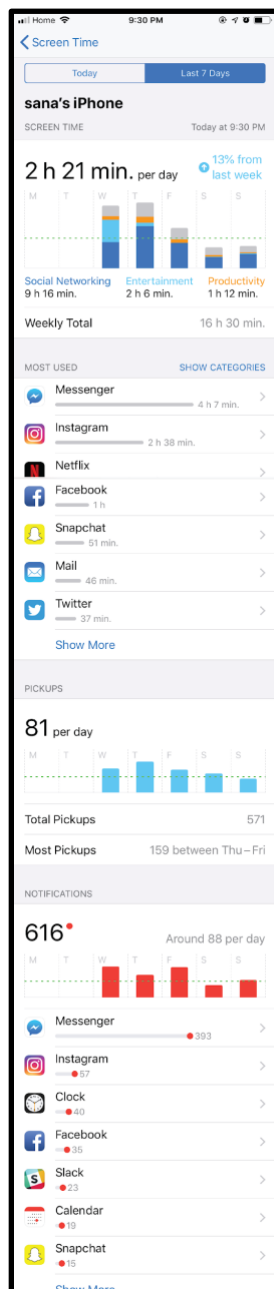


Figure 29. November 11, 2018.

This week, looking at my data it seems as though I had slightly less time on screen than I did the week previously; although I am a little confused why the app told me I was up 13% from last week when I actually had less time on my phone per average than the week finishing on November 4.

It also seems that I have less notifications this week, although of course Messenger is the constant in all of this. I had an interesting/terrifying dream the morning before I screenshotted this data; Zoya and I had taken a flight to Germany (makes sense because I was recently in Berlin) and the whole flight there I had a deep sense of unease and wasn't sure why. There were some people on that flight around us that just didn't seem to be what they claimed to be. Turns out that in my dream those people actually were part of a large con scheme which culminated, somehow, in the broken logic narrative of dreams, in the kidnapping of Zoya and I and accidental murder of Zoya!! I woke up in such anxiety and fear because in the dream I only witnessed her death and had to escape and never found her body or resolution or confirmation of what I had seen and that terror over not having the confirmation to know where she was was so painful.

There's some pretty obvious connections between having a dream like that and acknowledging that a distance from my family means that to some extent I am not able to be there for them physically, and whether that means not being able to protect them, or even just to be aware of where they are, it's something that I've had to learn and figure out this year that I've been away. It also points to the

state of my own anxiety, where not only am I anxious about my own life and getting all the stuff that I have to do out of the way and completed, but also that I am at the same time missing my home and/or worried about my family

November 18, 2018

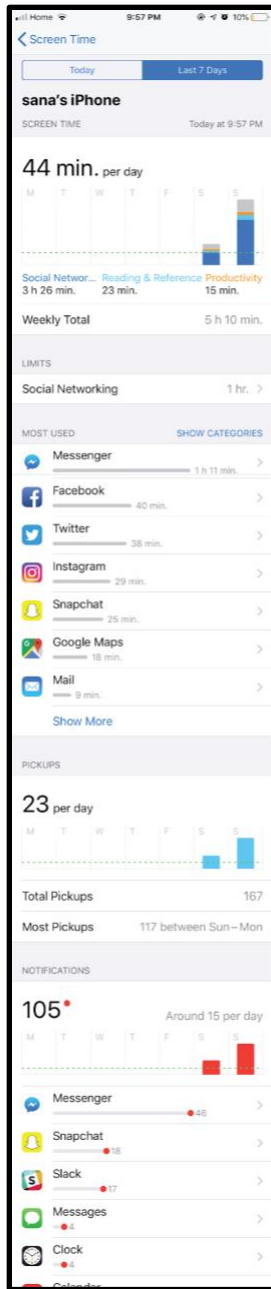


Figure 30. November 18, 2018.

I think I broke my screentime function. I think it might be because I kept capping my screen time allowance with certain apps, and for some reason that affected my phone's ability to track how much time I spend using it. I was trying really hard to be productive and enacting all of these limits on my usage of certain websites on my computer browsers, and time limits on social media apps on my phone... to be honest this doesn't seem to particularly help as I will inevitably just turn off the time limit and keep using those apps regardless. I seek out those distractions especially when I feel panicked about being productive.

November 25, 2018

December 2, 2018

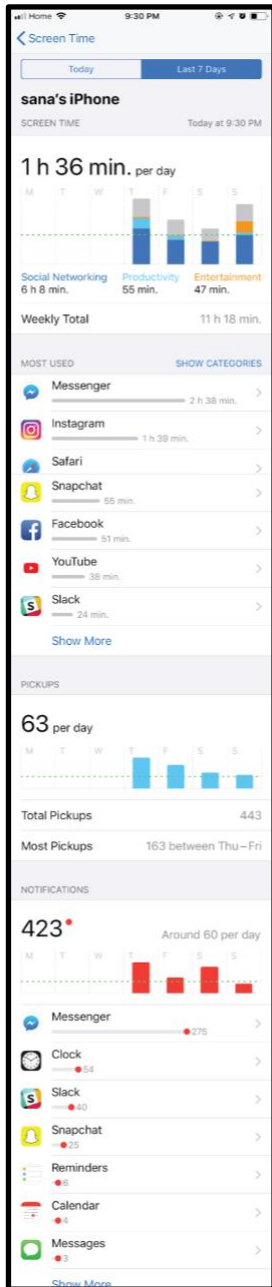


Figure 31. November 25, 2018.

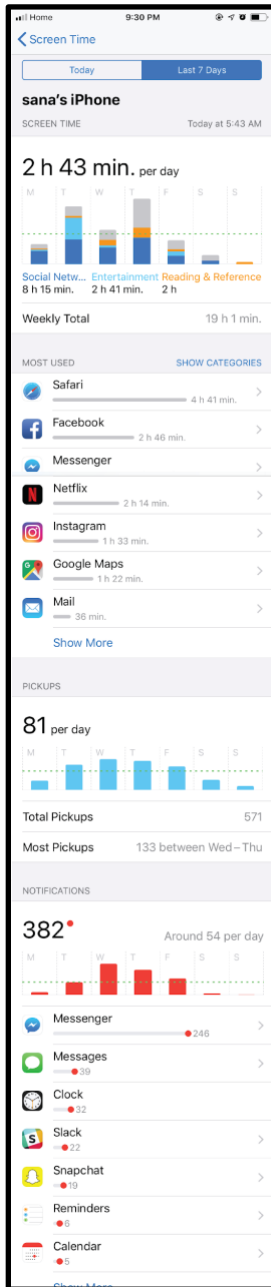


Figure 32. December 2, 2018.

It was my birthday this week, and my boyfriend and sister threw me a surprise birthday!

December 9, 2018

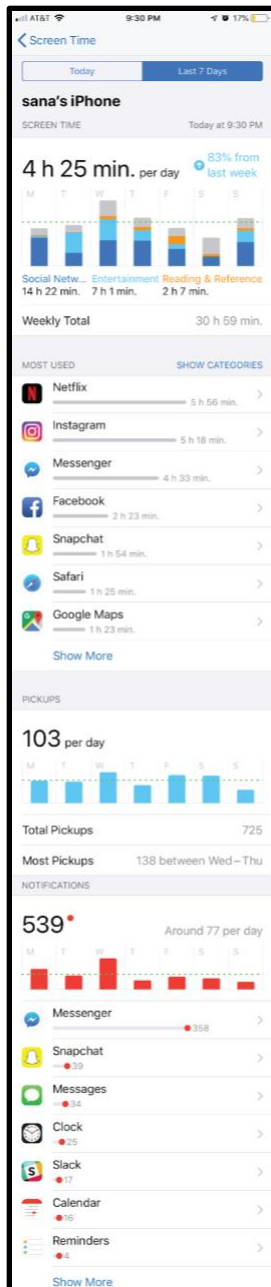


Figure 33. December 9, 2018.

(I keep trying to write about the internet but I keep getting entrapped within the internet. Between the moving images and memes I just get so lost.)

The date of these screenshots marks the beginning of my two weeks at home in New Paltz, New York, on my winter break, and the end of the fall semester. I did a fair bit of traveling and waiting around in airports so the 5 hours of Netflix does make sense. God, the 5 hours of Instagram is definitely not healthy though.

December 16, 2018

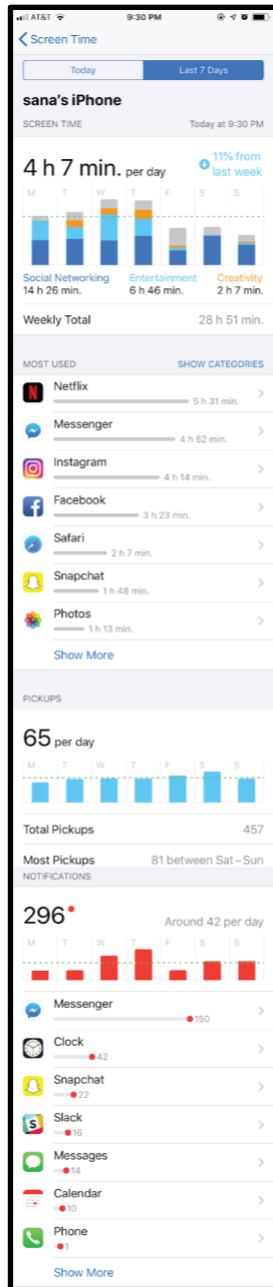


Figure 34. December 16, 2018.

I watched a movie on Netflix recently. It's a French film directed by Fred Cavayé, called *Nothing to Hide*. I assume it was recommended to me because as I recall it didn't take very long to have it pop up on the app's homescreen. It was about a group of old friends, 3 couples, getting together for a dinner party and deciding to play a game where all of the messages and calls coming in on their phones had to be read out loud, or taken on speakerphone. Essentially, all of the correspondences that are normally private had to be shared publicly in this dinner party setting. A big part of this highlighted infidelity amongst married couples, while another aspect that I found intriguing and something that could have been developed further in the film was a discussion of an intimate and private relationship with a personal device such as a smartphone. The scene where the game is first suggested demonstrates extreme unwillingness to part with these physical objects, probably because of all of the private content stored within it; and yet, as many of us know, the data contained within our devices is not private regardless of whether or not we have decided to share it amongst our immediate community. People are possessive of the devices themselves. Also, I notice even while writing this I have this urge to turn on and check my phone every couple of minutes. I know that I haven't received any notifications since my phone would vibrate if I did; and yet, if there's any moment where I require my brain to be working a little bit more focused than normal I put off the moment where I actually have to work by being like 'ok time for a quick phone break and then I'll write about all the complicated stuff.' Sorry, this is super non-academic but hopefully I'll

edit it to be more coherent later. By the way! I'm home right now and actually spending a lot of time with Zoya and the rest of my family in person which I honestly haven't gotten a chance to do in way too many months. The last time I was home for longer than a day was in April, 7 months ago. I honestly forgot how much I need to be here, and how healthy it is for me to be with the people who raised me and who I grew up with. It's not been a year and 4 months since I moved to Toronto, and I think I've become somewhat

accustomed to living there and being this different self than I am at home; but again, it really is a completely different way of living. Toronto is faster, more sociable, than a sleepy small town in the mountains in upstate New York, and yet at home I feel really understood. Here I feel completely in place. My sister still is working full time, though, so I actually don't think we've been messaging any less than we usually do during the day. I'll still send her random messages while she's out of the house, although again really nothing can replace being together in person. The messages are just little blips of updates, and they are not at all a substitute for real-time conversation. These sentiments are really reinforced in Sherry Turkle's "Reclaiming Conversation," where she discusses the phenomena of technology taking away from people's quality of time and conversation together. Students, going out to dinner with their friends, will have their phones on the table and don't consider it rude to be on their phones as long as there are at least 3 people carrying on the conversation. In this example, Turkle writes that the most commonly heard phrase at these get-togethers is, "wait, what?" when someone picks up their head from their phone and tries to jump into the in-person conversation without having followed the previous narrative. As a young person having gone out with people my age, I can attest that this is definitely something that happens. I myself am a frequent utterer of "wait, what?" in many conversations, and it is indubitably because I had been distracted by some other thread of conversation or interaction on my phone. Speaking of hanging out with my sister in person this past week, I also find it humorous that, while I'm thinking about this thesis and me and my sister's relationship online, even when we hang out in person we sometimes fall into this habit of both being on our phones while being physically together. The internet beckons constantly, even when the people we actually want to spend time with are right in front of us. I will also note that being engaged in conversation or just consciously present with another person takes energy. It is so much easier to be on your phone, distracted and being fed content that doesn't really necessitate a lot of mental consciousness, than to think of what to say to another person in front of you. Zoya wakes up really early for work during the week and comes home after dark (the sun sets around 4:24pm this time of year in New Paltz, New York) and I know how tired she must feel. I don't hold it against her that for her to spend time with me sometimes means that it's easier to sit together and be on Instagram than to sit together and just talk. I think I'm also very open in a way that she's not always comfortable with, and a lot

of our interaction is sometimes just me talking to her and her telling me whether or not I'm being an idiot. She's a bit more private than I am.

December 23, 2018

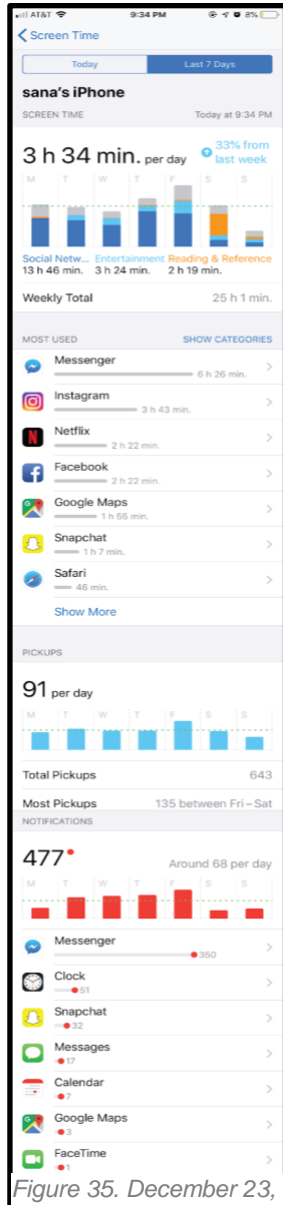


Figure 35. December 23, 2018.

December 31, 2018

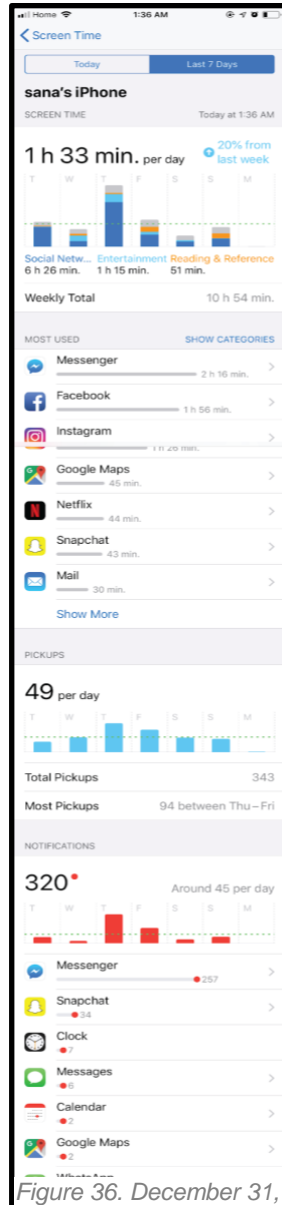


Figure 36. December 31, 2018.

Weird to see WhatsApp here, but pretty sure it was simply spam.

Note about this week: we had a very intense weekend of travel, rehearsals, and performances. Me and Zoya both perform with a Ukrainian group called the Women's Bandura Ensemble of North America, and this week/weekend was a whole mess of activity, which sort of explains my lack of usage of my phone.

January 6, 2019

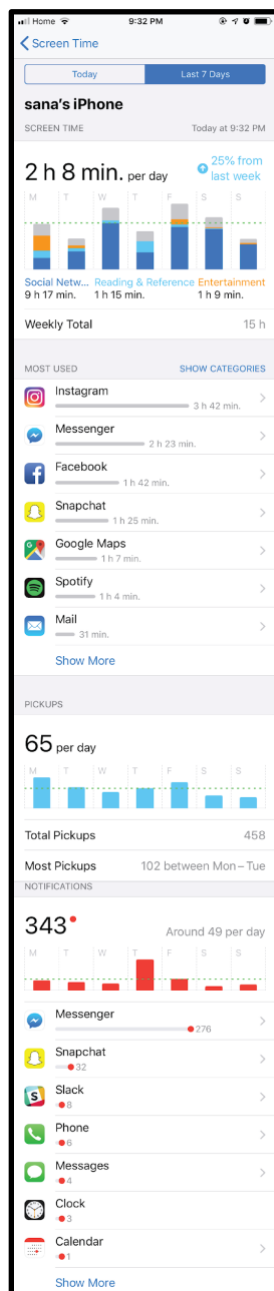


Figure 37. January 6, 2019.

Through this thesis I've become more and more interested in self-produced data online, and have gotten into the habit of requesting a copy of my personal data on any app or website that offers it. One thing that has come out of the monetization of personal information and its preceding public backlash is that there is seemingly some more transparency in regards to the types of data that is collected on people. However, I do sense that there is a literacy issue when it comes to data, especially in the format that most sites offer it in. Anyway, I just downloaded my Spotify data. That is an app that I truly adore. As someone who grew up playing and being surrounded by music, I have a really soft spot for anything that allows me to listen to it. My dad introduced Spotify to me and my siblings back when it was an invite-only, totally free streaming platform, back in 2008, and it totally changed the game for accessing music. My method for finding music pre-Spotify was finding songs on iTunes or an illegal website (shh) and downloading them to my computer. The concept of streaming wasn't that big, UNTIL Spotify. Going through that dataset and seeing the music I listened to, the searches I made, the playlists I created, is just as nostalgic for me (if not more) than going through my Facebook data. This is mentioned time and time again, but the performance of social media separates the essence of a person from the digital self they project. The other day, I was mindlessly consuming Instagram stories (which recently came out with a feature where the story instantly moves onto the next without needing any interaction from the viewer, basically making Instagram stories into its own self-produced, collaborative Netflix movie) and one guy's story stood out to me. I know him from undergrad, back in New York State.

It read "so do u not love digital me?? Or is it just insta gross algorithm that got no one paying attention to me? Anywho I'm better irl... want to have more social physical realness in 2019 so hmu let's have a sleepover" the next part of his story is a mirror selfie in what appears to be a bathroom, of his shoe in a

sink. Also, just sometimes humorous things stand out to me as I go through the dataset containing messages between me and Zoya; for example, in 2017 I sent her 'goddammit' 6 times.

January 13, 2019

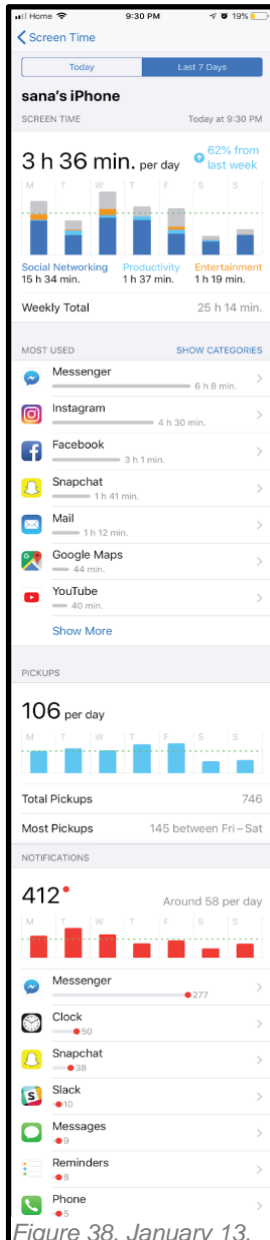


Figure 38. January 13, 2019.

A lonely week, as told by my data.

I downloaded a fresh copy of my Facebook data. Interestingly, there seemed to be issues opening the message folder downloaded from Facebook. Everything else I could open, except for that folder.

I reported the issue to Facebook but was struck with how little customer service representation the company provides. There is no way to get in touch with any actual Facebook representative, not even an AI bot. After submitting the report message, it even notifies you not to expect any actual response.

January 20, 2019

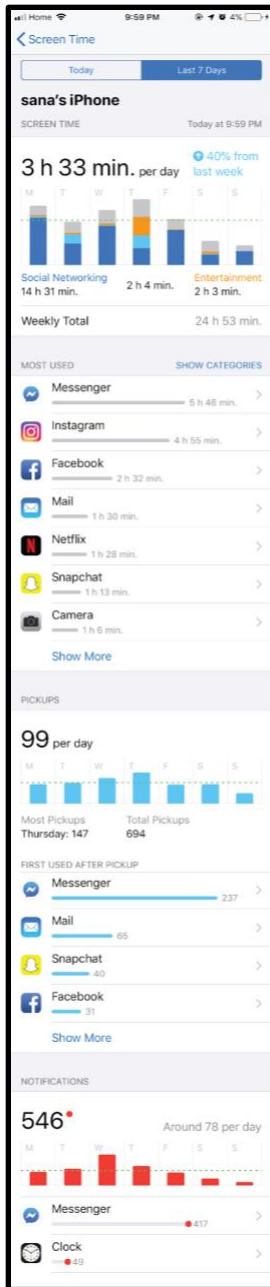


Figure 39. January 20, 2019.

This was also a busy week of performances and rehearsals, but my usage screenshots are telling, because I also spent a lot of this week alone. This time we were performing at a dance event, where our band performed. Zoya and the rest of the band arrived late on Friday night, and we spent all of Saturday rehearsing and Sunday they drove back home.

This week I noticed myself becoming incredibly distracted and losing a lot of time in spending hours on various social media, both on my laptop and on the apps on my phone. It was a lot of mindless scrolling, and I could feel that this wasn't feeling particularly good for me, so I decided to try to cut it out as much as possible. however, now I feel incredibly lonely.

January 27, 2019 February 3, 2019 February 10, 2019 February 17, 2019 February 24, 2019

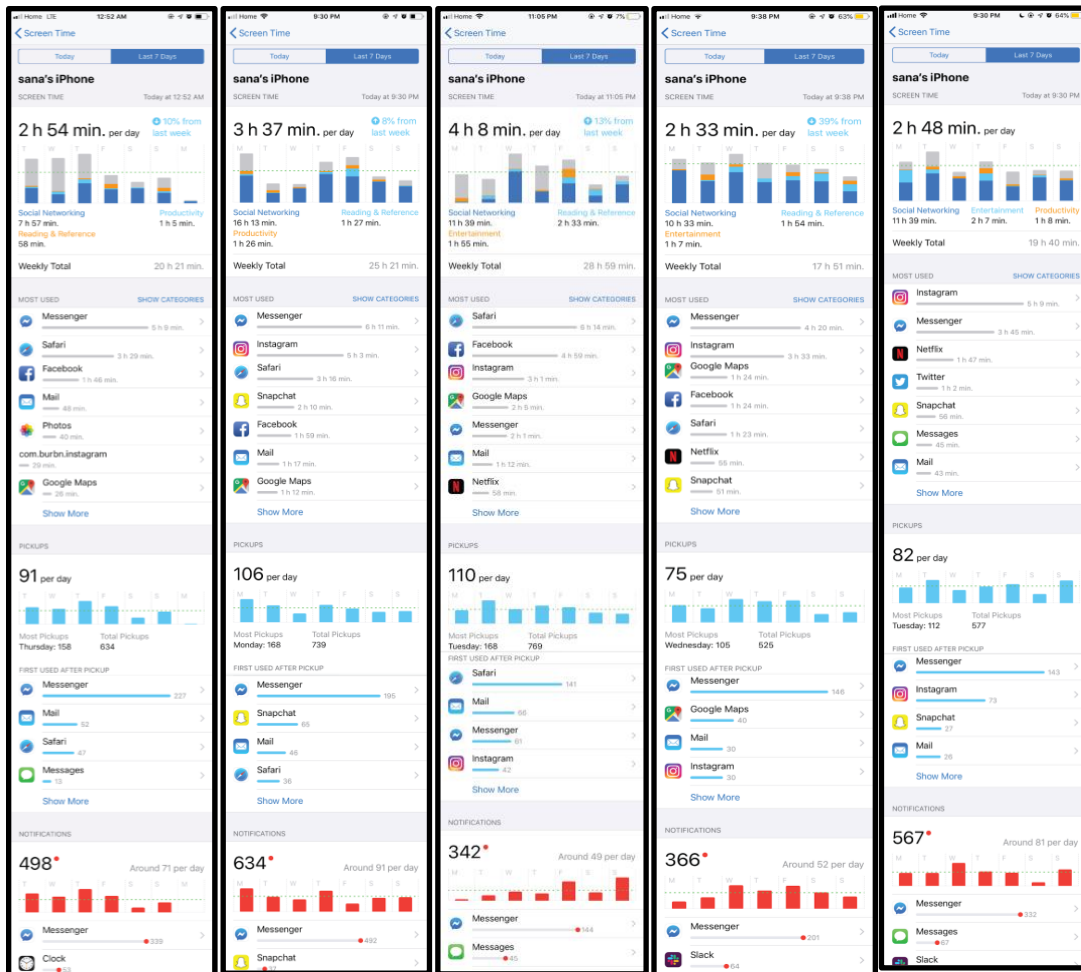


Figure 40. 1. Figure 41. 2. Figure 42. 3 Figure 43. 4 Figure 44. 5

Screen Time - Conclusion

The collection of the Screen Time usage shows that I spend a lot of time on my phone during some weeks, but more importantly it validates choosing Facebook Messenger as the social media data to visualize. Screen Time shows that Messenger is the app I spend the most time on, as well as the app I receive the most notifications from—indicating that it is a highly interactive social media. This is important, as this means the data from Messenger is likely to be insightful and the most expressive in visualizing digital relationships.

Appendix B - Accessible charts

The data visualizations created for this thesis relied heavily on colour in order to display different information. At first, colour was commonly used in the visualization to denote the sender; however, in later prototypes colour was used to show the 11 different sentiment groups. The choice to use colours to show emotional groupings was based on research which studied human association of certain colours with emotions. As such, the information shown through the visualization becomes inaccessible in instances when colour is unavailable. This section contains all of the data visualizations from the body of the thesis document, redesigned in black and white.

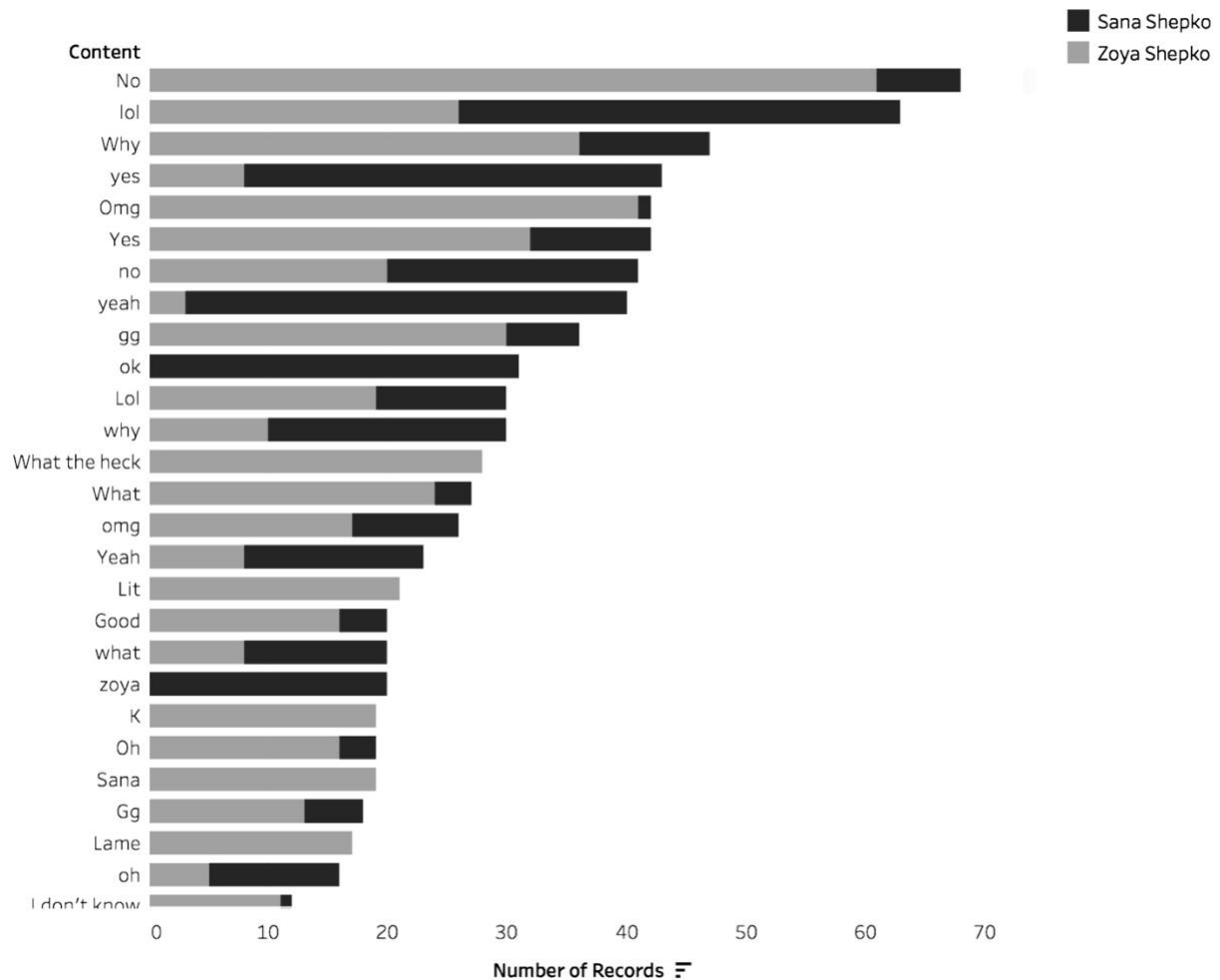


Figure 45. Black and white version of Figure 6.

FACEBOOK MESSAGES SENT BY SENDER (color) AND NUMBER OF RECORDS (size of circle)

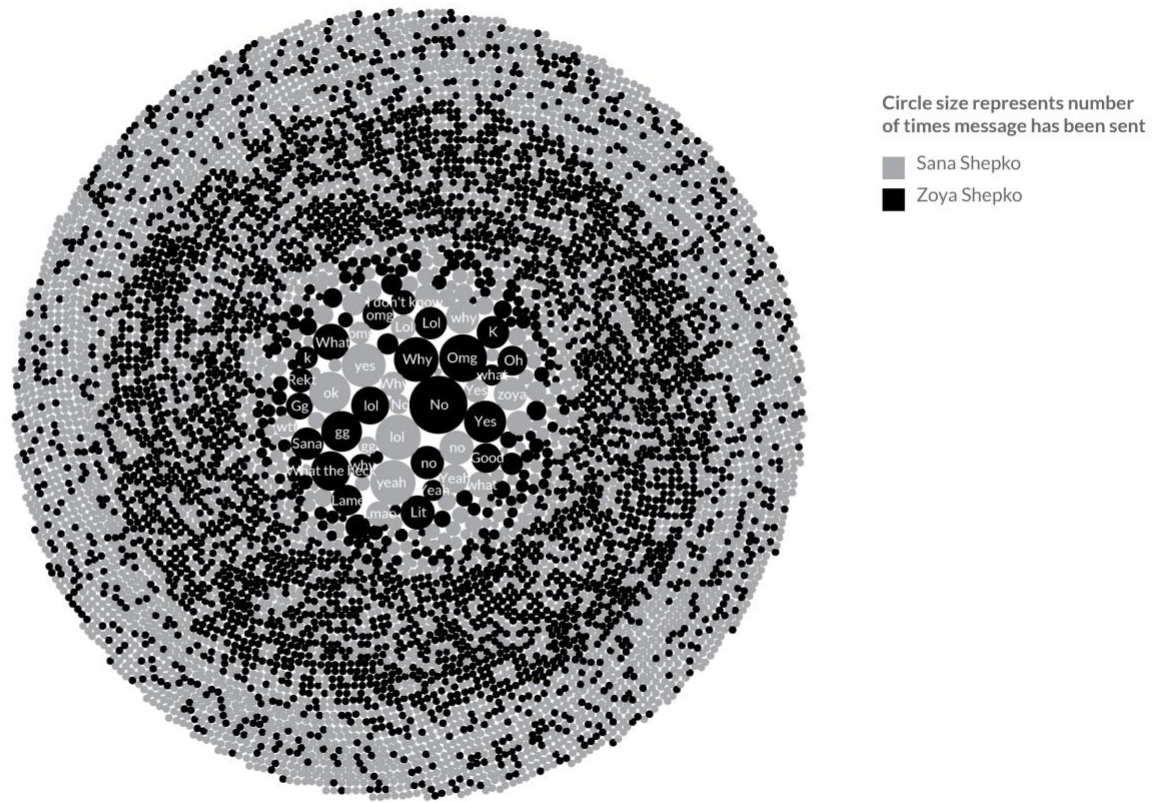


Figure 46. Black and white version of Figure 7.

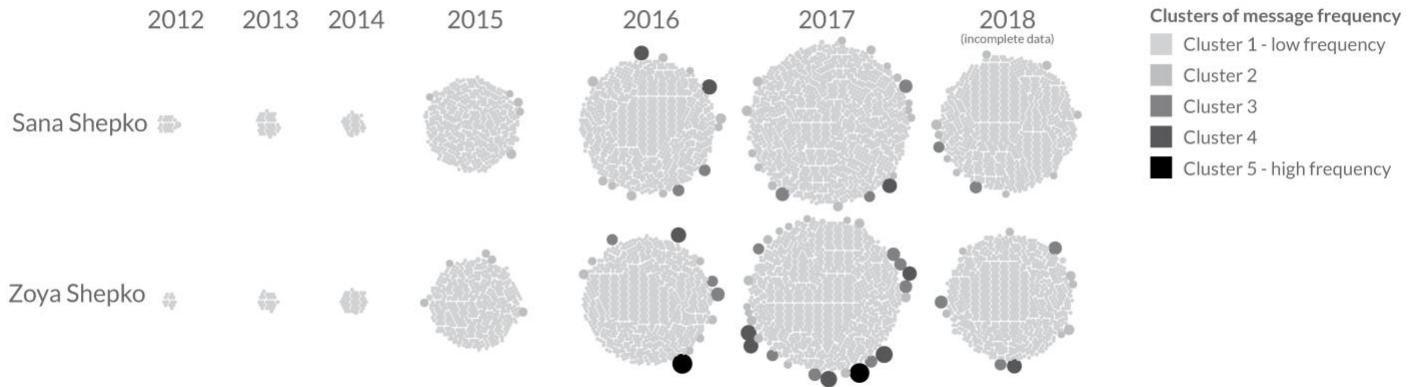


Figure 47. Black and white redesign of Figure 8.

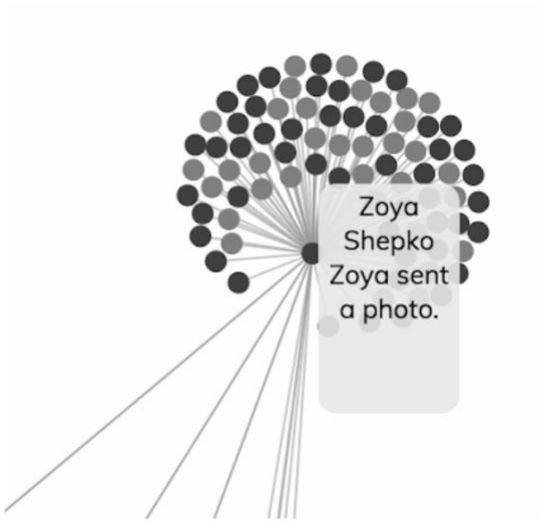


Figure 48. Black and white version of Figure 13.

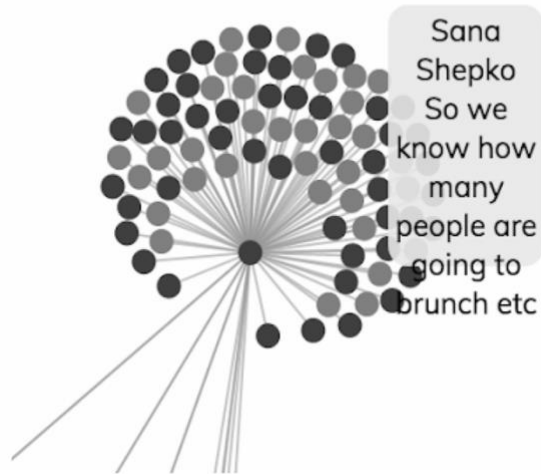



Figure 49. Black and white version of Figure 14.

Sana Shepko

my sister in 

network

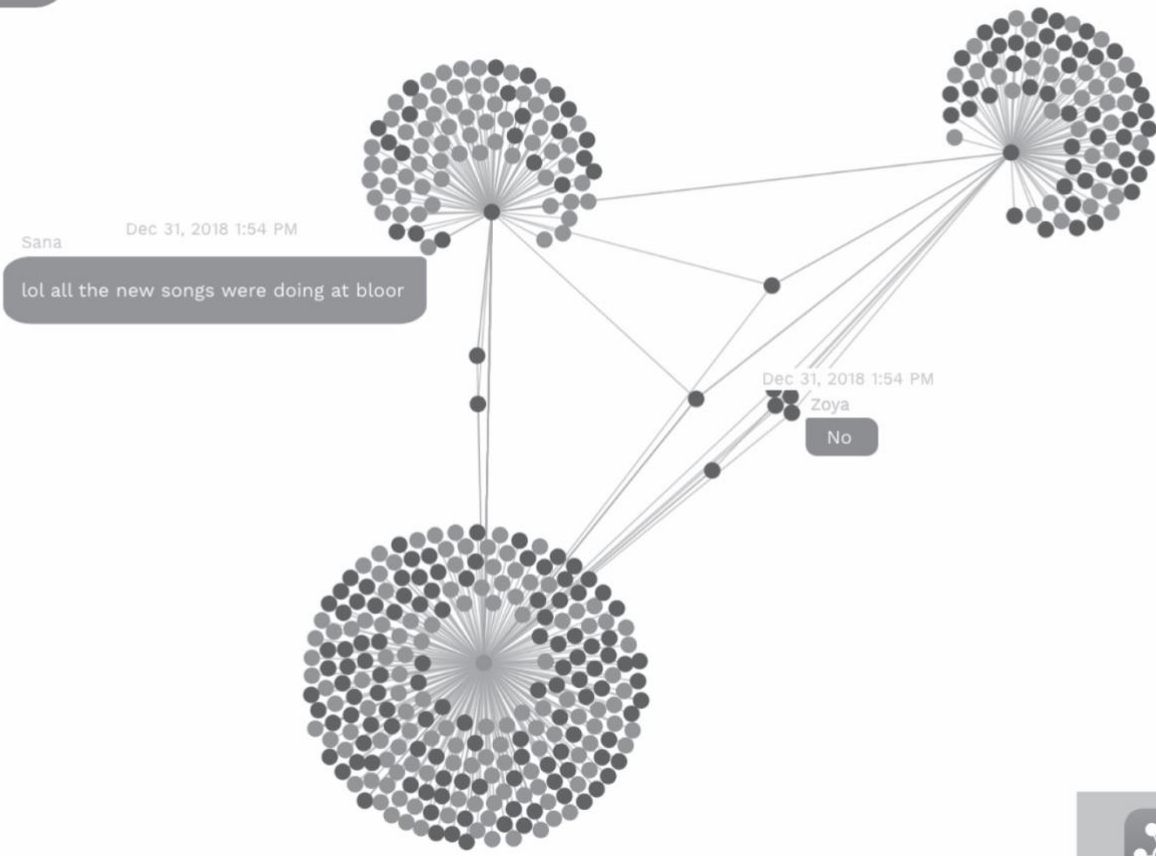


Figure 50. Black and white version of Figure 15.

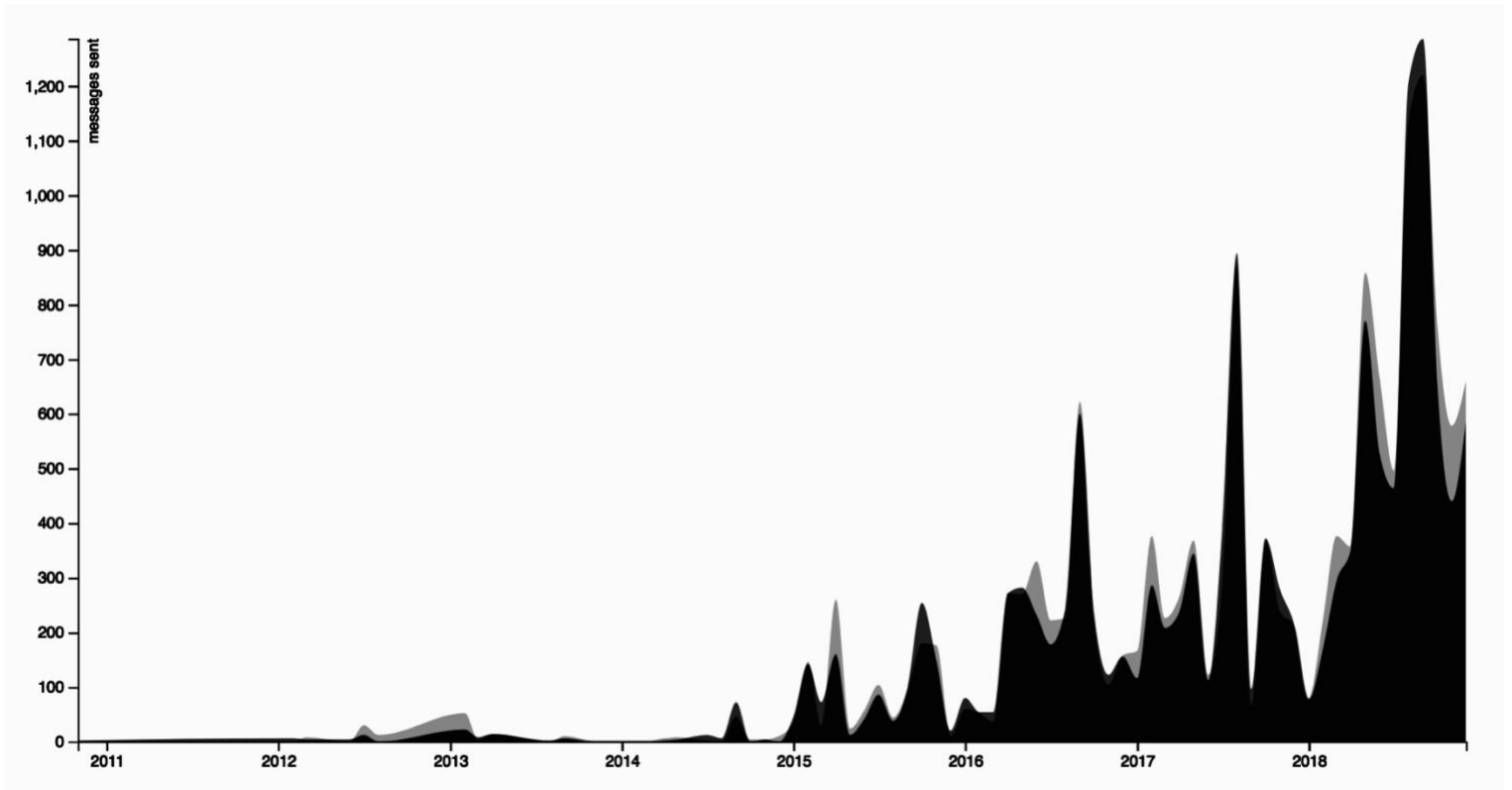


Figure 51. Black and white version of Figure 22.



Figure 52. Black and white version of Figure 24.

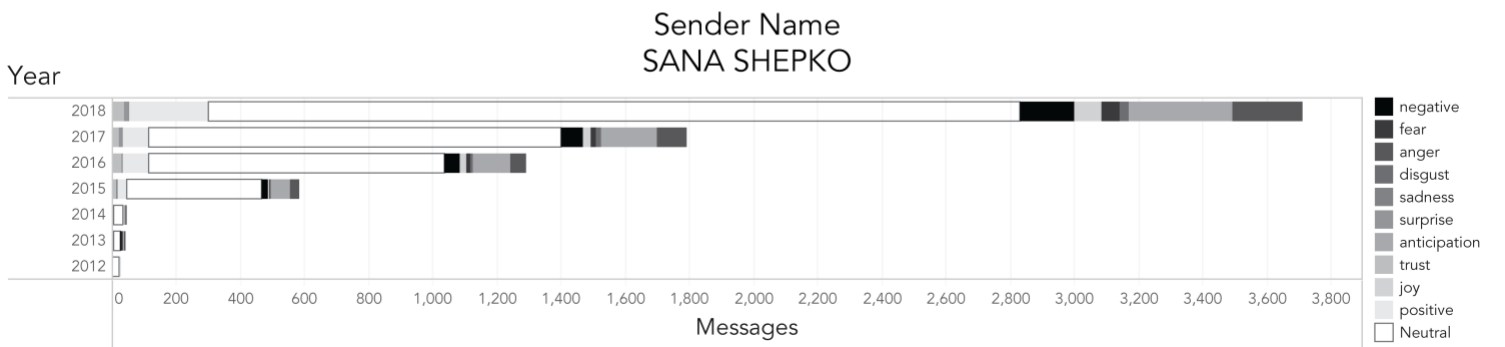
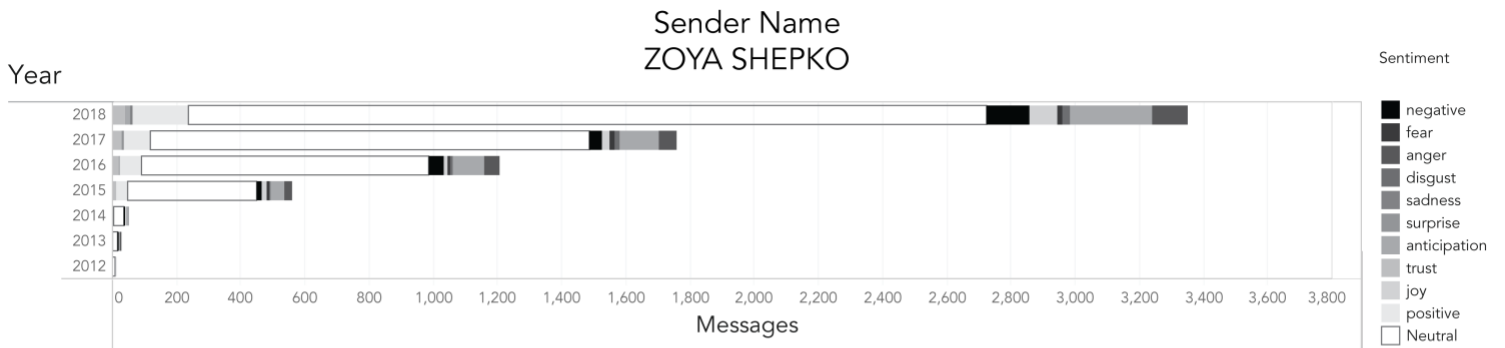
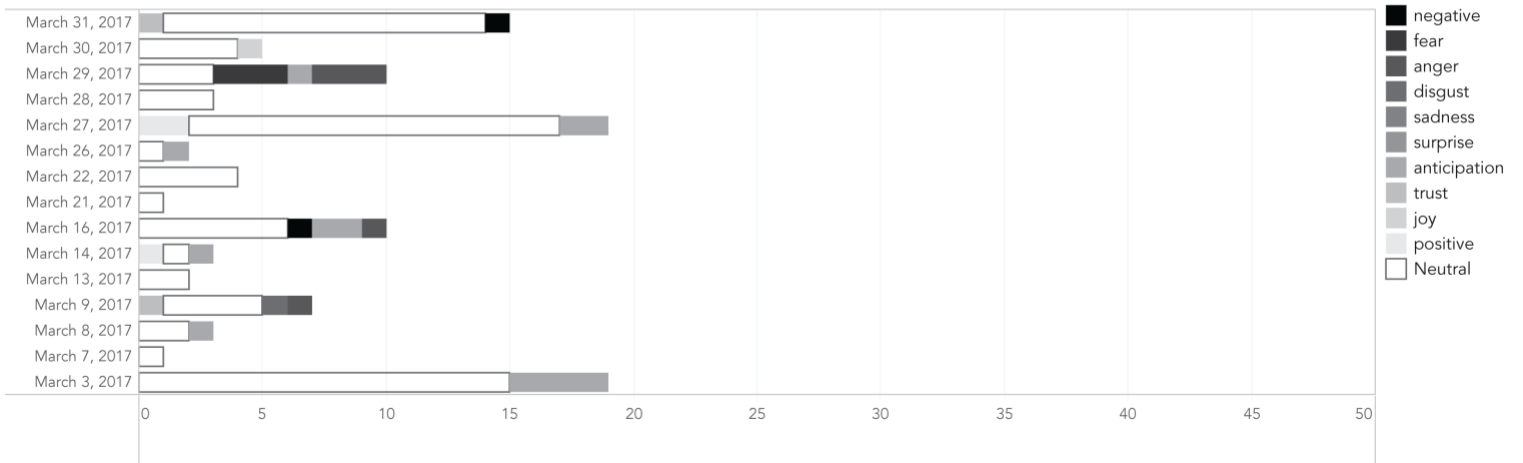


Figure 53. Black and white version of Figure 25.

Sender Name
ZOYA SHEPKO

Days March 2017



Days March 2018

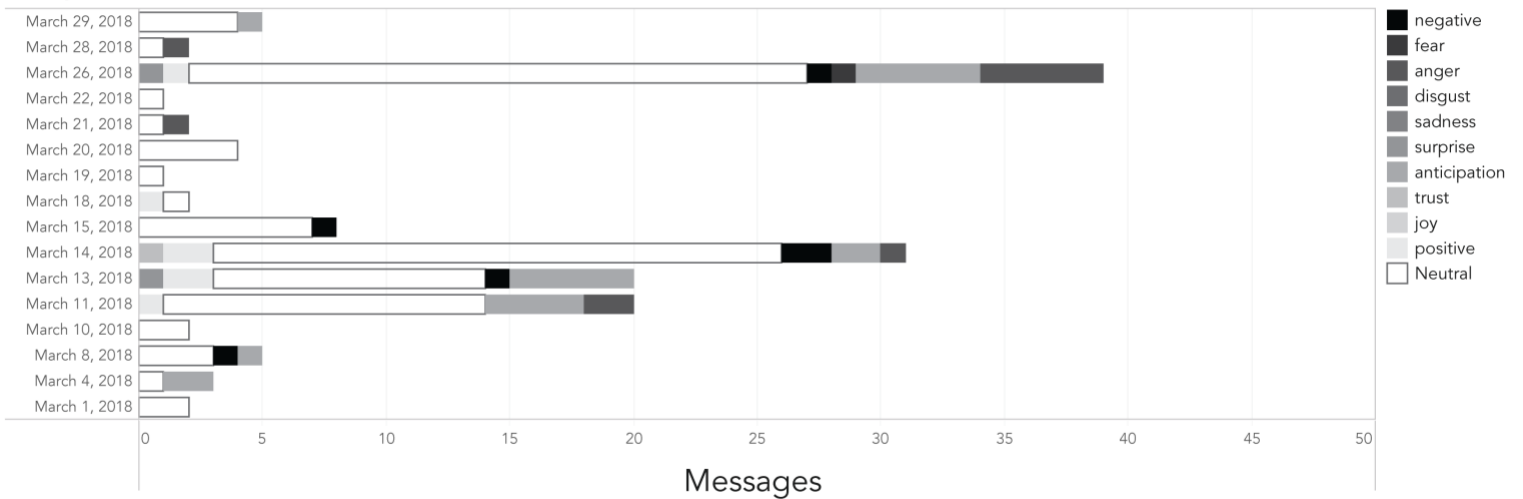
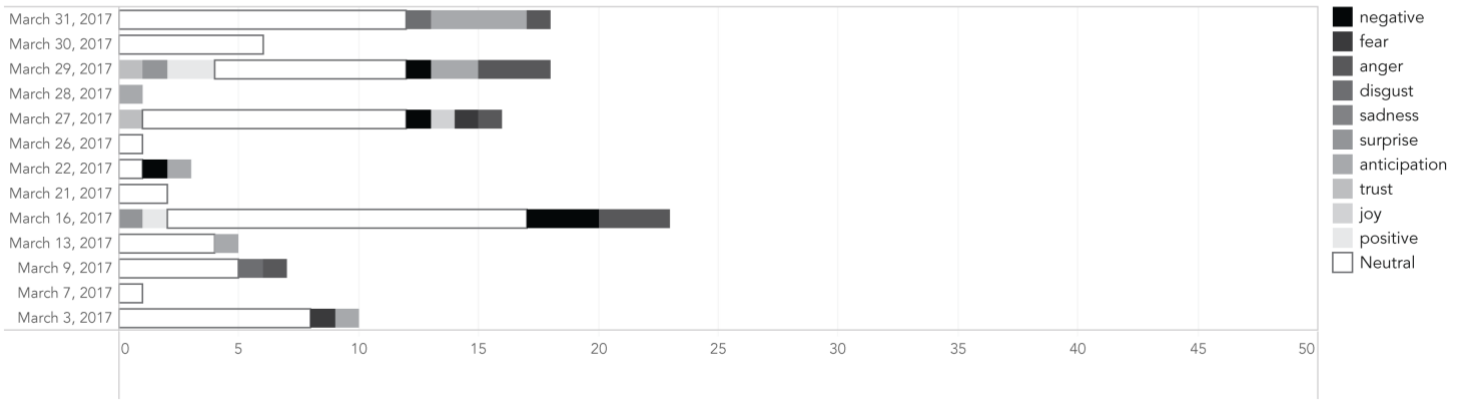


Figure 54. Black and white version of Figure 26.

Sender Name
SANA SHEPKO

Days March 2017



Days March 2018

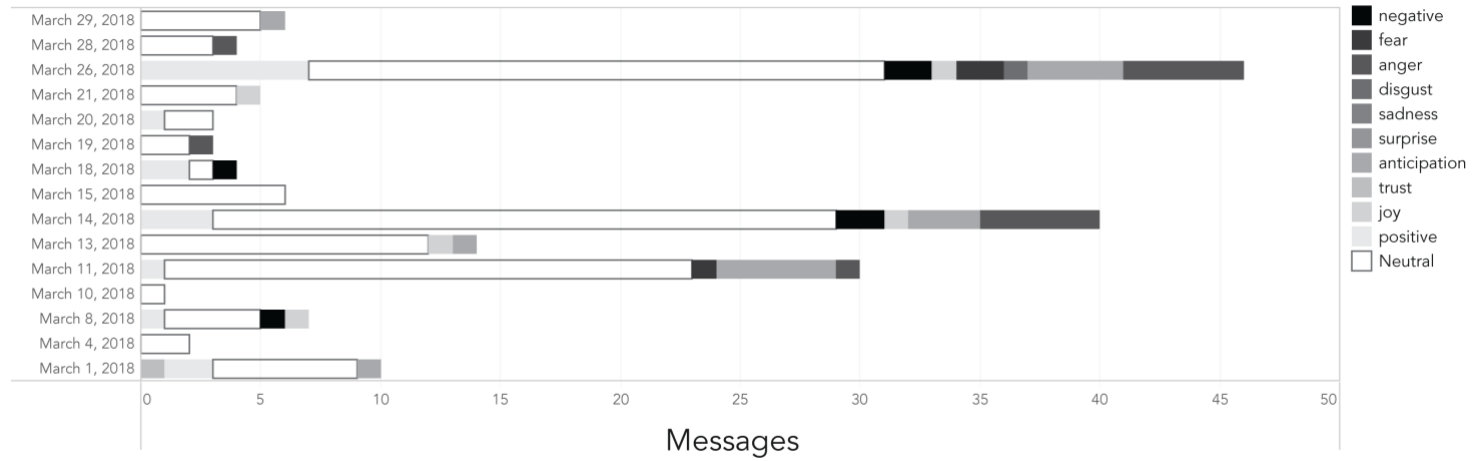


Figure 55. Black and white version of Figure 27.



messages from zoya

messages from sana



Figure 56. Black and white redesign of Figure 47.