TIED IN KNOTS: A CASE STUDY ON ANTHROPOGRAPHIC DATA VISUALIZATION ABOUT SEXUAL HARASSMENT IN THE ACADEMY

Abstract
With this pictorial we present the design process of “The Academia is Tied in Knots”, an interactive visualization based on sensitive and qualitative data, namely personal stories reported by people who have experienced sexual harassment in academia. We discuss how we approached the task of visualizing sensitive, uncomfortable, yet important topics in terms of data-mapping and visual representation, including the appropriateness of computational vs. manual approaches to help foreground relevant themes. We also describe the design process behind the visualization and we discuss it from a feminist data visualization perspective.

Introduction
Sexual harassment (SH) is a serious issue at workplaces across all sectors - and academia is no exception. Yet, SH in the academy has only recently started to become acknowledged and addressed. SH has a negative impact on individuals’ lives and careers but also affects the educational environment as a whole; it threatens academia as a place that, ideally, should create a safe and supporting environment that fosters knowledge generation and discussion, free of individual suppression. Although the issue of SH is widespread and has a damaging impact not only on academia but also on society at large, until recently, it has been little discussed or addressed in academic circles and at the institutional level. In late 2017, the anthropologist and academic advisor Karen Kelsky released an online survey with the aim of anonymously collecting the experiences of people who faced SH in the academy. These testimonials were collected in the form of semi-structured textual data [1]. Kelsky’s goal was to provide a platform for people to share their stories, highlight the extent and impact of SH on people working in academia, encourage solidarity and, ultimately, to inform change. The survey consisted of predominantly open-ended questions asking participants to describe their experience of SH in their own words, the academic context in which the incident occurred, and the impact the incident had on their own (academic) life as well as repercussions for the harasser (if any) - see the list of questions and an example testimonial in the figure at page 3. Within approximately one year, more than 2000 people participated in contributing their experiences with SH in academia, highlighting not only the spread of the issue across disciplines, institutions, countries, and gender, but also the subtle nuances of SH in academia and its (often severe) impact on the victims. The survey itself received a lot of media attention and inspired articles around the topic of SH in academia. Kelsky also made the survey data publicly available in the form of a spreadsheet [1]. While the individual stories stand for themselves, it became clear that this form of data representation did not do justice to the extent and richness of the data, nor did it provide a powerful entry point to raise awareness and promote discussions around the topic within academia and beyond. Inspired by current discussions on data feminism [2], ethical considerations of visualization [3], critical InfoVis [4], approaches to data visualization in the digital humanities [5], and anthropological approaches to visualization that focus...
Users can freely zoom and move around using pointer or touch devices. In addition, the zoom can be controlled with a designated slider.

By clicking or tapping, knots unravel and show further information.

Enable or disable voice records.

Switch between three positioning modes.

Use this filter to highlight elements with specific data categories.
on the people behind the data [6], we as visualization researchers and designers became interested in the following questions. (1) How can visualization help to give visibility to the issue of SH in academia? (2) When visualizing this type of data, how can we honor and empower the individual voices within this data in a sensitive way, rather than silencing them within abstract and aggregated views? (3) How can visualization promote sensitivity and awareness of the nuances of SH in a way that encourages constructive discussions of how to make academia a space that is safe for everyone?

These questions pose interesting visualization challenges such as how to create tailored visualizations that uniquely reflect on the given data set and represent people’s individual and deeply personal experiences behind this data. A second challenge is coming up with appropriate visual representations that are effective and understandable by a wide audience, yet act as a defamiliarization with concepts [3,4] that we often take for granted and provoke discussion.

In this pictorial we describe and reflect on our design process of visualizing the SH survey data which involved the interdisciplinary collaboration between researchers with a background in graphic design, visualization, human computer interaction, natural language processing, social sciences, and anthropology. We illustrated this process through unfinished visualizations [7], also described as “visualization sandcastles” [4], that we built as part of this process and that finally led to our web-based visualization “Tied in Knots” (see page 2) which - based on the metaphor of knotted threads - represents evocative statements by contributors to Kelsky’s survey in form of text and sound which, in turn, provides an entry point into survey participants’ full testimonials.

We start by describing our approach to familiarizing ourselves with the survey data and our subsequent data transformation and interpretation that we applied in preparation of our visualization process. This is followed by an overview of our iterative visualization design process, and a reflection on lessons learned in terms of visualization design methodology when it comes to visualizing qualitative sensitive data - see figure above for an overview of our design process.

Data Exploration, Interpretation, Transformation

Sampled Close Reading.

Our first step was getting to know the data in the search for potential angles that our visualization could provide on these personal...
Describe the incident(s) [free text field]

«When I was an undergraduate student in my final year in 2005, I was groped and kissed by a professor while in his office during a private meeting.»

What was your status when the incident(s) happened [free text field]
«I was an undergraduate student in his class»

What Was the Status of the Perpetrator(s) (Particularly, relative to you)? [free text field]
«He was the professor of a class I was taking. He was the thesis supervisor of my roommate.»

What Was the Gender of the Harasser? [select from certain options or specify in a text field]
«Male»

What type of institution was it? [select from certain options or elaborate more in last field]
«Elite Institution/Ivy League»

[Optional] What Was The Name of the Institution(s) [free text field]
«McGill University»

Your Field/Discipline [free text field]
«Communications»

Institutional Responses to the Harassment (If Any) [free text field]
«None. Unreported»

Institutional/Career Consequences for the Harasser (If Any) [free text field]
«N/A»

The Impact of the Harassment on Your Career [free text field]
«I currently work at the University where this harassment took place. I have taken steps to avoid encountering the harasser.»

The Impact of the Harassment on Your Mental Health [free text field]
«Distrust of authority figures. Depression.»

The Impact of the Harassment on Your Life Choices/Trajectory [free text field]
«Unsure.»

Other Comments You’d Like to Add [...] [free text field]
«Thank you so much for doing this. When the #MeToo movement started, I finally told my friends and family about what happened. I am still too afraid to report the harassment at the University, even though I know that this man has most likely continued to harass women in his classes. I felt that no one cared about what happens in the academic environment. I feel it is one of the most patriarchal environments to study and work in.»

testimonials. For this reason, we engaged in close-reading of samples of these testimonials. Close reading in preparation for visualization can be considered a rather unusual approach to visualizing data, and it is better known from visualization projects in the context of humanities research. This approach helped us to gain a deeper understanding of the experiences of the victims of SH in academia and the character of the survey data at a low level; the whole research team engaged in these investigations. The close reading approach revealed the intimate and detailed nature of the testimonials, that often describe victims’ personal feelings associated with the incident. It also revealed the varied nature of incidents of SH (e.g. verbal or physical), the characteristics of perpetrators (often males higher up in the academic hierarchy, often in the role of supervisors), and the range of contexts in which incidents happened: private offices, off-campus bars, conferences and social events.

The close reading approach was invaluable in that it provided all team members an in-depth understanding of the overall qualitative nature of the survey data and its (often disturbing) details. However, we also felt the need to incorporate computational methods to explore the data in order to gain a higher-level understanding of survey responses as a whole.

Left: Survey questions and answers taken from report #1055. Data retrievable from: https://bit.ly/survey-sh-academy

Data formats sh-academy
Distant Reading: Looking for Language Patterns.

To complement close reading, we applied common computational text-mining approaches to extract structured information from the open-ended answers of survey participants. Named Entity Recognition techniques were not used because all references to individual identities had been removed due to privacy policies and because names of places or institutions were rare or unreliable. Instead, we focused on identifying patterns of language usage and recurring phrases within the dataset. Reports were split and regrouped by survey questions and the texts were fed into a Python script that used NLTK to extract and count n-grams. For the sake of experimentation, we tested n-grams of different sizes and different text pre-processing options: for instance, texts have been used as-they-are and with stop-words or punctuation removed. The extracted data was then used to produce treemap-like visualizations in RAWGraphs [8], that were then exported as SVGs, reworked and annotated using Adobe Illustrator and exported once more as SVG files in order to enable the quick addition of interactive elements to the SVGs with web programming languages: HTML, CSS and Javascript. Annotation was key to help reflection on these quickly generated and transient visualization prototypes that acted as mediators of discussions within the interdisciplinary visualization team. As such, they were never intended as communicative artifacts for the general public, but as intermediate and analytical milestones in an iterative process of research. However, they quickly allowed us to grasp some interesting aspects of the data that eventually found confirmation in what we had learned from the close reading of sample testimonials.

Our computational text analysis revealed “I was” as the most common bigram across all of the testimonials, used 2,647 times in the incidents descriptions. This fact highlights the deeply personal and situated nature of the data. Interestingly, “He was” is the most frequent bigram in the descriptions of perpetrators. The phenomenon of SH appears to be bound to a verbal dimension, and less to a physical one; however, this does not render the described SH incidents as less harmful - anxiety, depression and other mental health issues stand out from the visualizations as consequences of the SH for victims.

Our computational text analysis also confirms our impression from the close reading of testimonials: SH is closely related to hierarchy and harassers are most frequently described as “full professor”, victims are predominantly described as “students”. In terms of consequences for harassers, there is a strong dichotomy between “Title IX” (a law in the USA in defence of civil rights) and ‘no consequences’. Moreover, many survey participants admit to not having reported the SH incident that occurred to them. By digging...
voices and underlying emotions within their statements. We felt that re-grouping statements based on the results of our text mining approach also hampered the reading experience and was not empowering for the community at all. Reflecting on the generated visualizations themselves, we felt that the treemaps appeared aseptic in the context of this sensitive data set and did not convey the deeply personal and emotional nature of the survey data. However, this particular visual experimentation allowed us to enhance our understanding of the data as a whole and we consider it as a first experimental milestone of the longer process of visually representing this complex data set. This milestone led us to explore other methods of data transformation and metadata extraction, as well as less conventional methods of visualization. Reflecting on the problem, it became clear that we were in the need of different methods and different visual languages capable of treating this material in a way that was both respectful to the people who created it and that involved the members of the academy in contemplation, reflection and self-criticism. After reflections and discussions, we decided to apply a qualitative coding approach and collect metadata from a higher level set of tags. These tags are representative of certain traits of the survey entries that we came to know thanks to close and distant reading and that we considered as the most important aspects of the testimonies. We made this choice on top of our own understanding of the survey and in this sense we consider this a curatorial approach to the collection of this metadata. The idea that data is purely objective has been already challenged in the field of visualization, for example in the digital humanities [5] and more recently by the feminist movement [2]: data is not ‘the truth’, it may be subjected to political powers and it may perpetrate idiosyncrasies or imbalances against minorities. We decided to lean on this concept and to not fear the idea that the data may embed subjective aspects, provided they can help in bringing to light the most contemplative and disturbing aspects of SH in the academy. Certainly, this process brings into play many decisions that are arbitrary and
The Top 50 **Bigrams** (raw text)

Represented in the form of a broken-down treemap (top 10 highlighted), where related fields are clustered together and titled.

No case transformation, stopwords are preserved, you can pan and zoom around. [Download viz](https://vialab.github.io/i-dropped/top-50-2-ngrams.html)

Above: one of the visualizations of language pattern extracted with text-mining used as example

Explore the interactive visualization at:

[vlalab.github.io/i-dropped/top-50-2-ngrams.html](https://vlalab.github.io/i-dropped/top-50-2-ngrams.html)
hardly possible to be reproduced, but the set goals and the explorations described so far, led to the decision that, at least in this specific context, data visualization practices might be less oriented toward the scientific tradition from which they came and could be used in a speculative and political way.

Open coding
The most relevant information that we managed to identify concerns the description of incidents and the status of people involved, being victims or perpetrators. Such information is very difficult to automatically extract from the unstructured text of the survey and, due to this reason, it was necessary to manually code the data by closely reading individual entries and tagging them according to the coding schema above. As the data was heterogeneous, for certain reports it was not possible to apply all tag categories. Instead of forcing imprecise data values we decided to leave some cells empty, with the idea of taking this aspect into account in the visual representation. The coding also entails a selection of excerpts from the texts, that is based on these criteria: it corresponds to a behavior that negatively affects academia and that ought to be avoided (pedagogical choice), it is heartbreaking (rhetorical choice) and it helps in remembering the story and in differentiating it from the others.

This process of data harvesting is very time-consuming. For this reason, only 10% of the data has been coded so far, and the task is still ongoing. The team deliberately avoided crowdsourcing services that hire workers remotely (like Amazon Mechanical Turk) because of ethical concerns: it would have been too difficult to guarantee appropriate psychophysical working conditions for those people asked to handle such sensitive and disheartening material.

Visual representation
After having harvested and structured the data, we moved on the problem of visual representation. Two main design requirements emerged from the reflections on previous steps. First, the visualization of data ought to be engaging and capable of inducing reflection and self-criticism in the academic community. Instead of focusing on analytical visual solutions, we then explored the idea of a visualization that invites contemplation and open-ended explorations. It immediately appeared as promising the idea of a digital space where we could organize data and let users perform their own path of exploration driven by curiosity and surprise.

Second, we wanted to ‘honor the stories of the survey participants’. This specific visualization problem, after a series of iterations, turned into the pursuit of a form of representation that entailed emotional dimensions, allowed for self-identification and was capable of bringing the personal stories in the foreground.

The first design requirement has been addressed by using a statistical method to compress the multiple dimensions of the
data into bidimensional vector spaces, where entries with similar features get positioned near each other. A total of three spatializations were created: a first one depending on incident descriptions only, plus two others, specifically conceived to create self-identification, that position entries according to victims’ and harassers’ status. In this way the viewers of the visualization have a chance to confront themselves with incidents experienced or caused by someone in a similar position.

The second design requirement was more complicated and required researchers to undertake a longer iterative process. A first attempt was oriented towards the possibility of representing all texts without reductions, insisting on commonalities and intersections. This strategy was abandoned, because it was evaluated to be too chaotic, as could be expected. However, we learned from our previous attempts that it was necessary to produce a reversible operation capable of reducing the amount of displayed text that did not break up the stories into inexpressive fragments. It is due to this reason that we decided to collect report excerpts. Afterwards, we explored different solutions for their representation, including bidimensional and three-dimensional springs and spirals. Eventually, our decision fell on the use of bidimensional knots. During this part of the
The Academia is Tied in Knots

Knots proved to be a convenient solution that works on a number of different levels. First, the shape of knots can be parametric, and so they can be used to encode data in an unusual and curious new way. Second, they can be used rhetorically to great effect: “I have a knot in my stomach” is a very common colloquialism (used in at least one account) that expresses uncomfortable feelings similar to the ones experienced while reading the survey. Thanks to knots we can create a visual connection between testimonies and the feelings they transmit. Another connection lies in the pronunciation of ‘knot’, that it is similar to ‘not’, the adverb of negation. We recognize this connection because negations and negativity are countless present in the survey, being so numerous the cases in which victims “did not report”, “hadn’t been heard”, decided to drop their career, lost self-confidence, etc.

Using D3.js, we automated the drawing of knots using information within the incident descriptions. Then individual elements were positioned in an abstract Cartesian space according to the coordinates of the vector spaces. Knots happen to overlap when they

---

**METAPHOR OF KNOTS**

- (The Academia is) tied in “not”s
- Knot > Not

---

1. Exclusivity
2. Distress
3. Tied in “not”s

---

Hours: I did not report
People did not continue career
Nothing happened to the harasser

---

“Sleeplessness; knot-in-stomach; general exhaustion;”

---

Top: explorations of designs for representing individual stories in form of text
Bottom: using the visual metaphor of knots to display stories excerpts
VisAP’20, Pictorials and annotated portfolios.

Testimony #38

"He hit on me repeatedly, which freaked me out b/c he was much older and had all the power."

Incident description
- Verbal abuse (14)
- Higher hierarchy (18)
- During business hours (17)
- Not at a conference (7)
- Incident not reported (5)
- [impossible to evaluate successfulness of report]
- Victim stayed in Academia (4)
- Victim avoided people or places (18)

Data mapping and control-points grid

Knots are created using a CatmullRom curve interpolation that uses as control-points centers of the enumerated circles of this grid, that represent data values (legend below)

In case certain information is not available, corresponding grid points can be skipped

Artificially added control points to improve stability in curve extremities

Parametric drawing of a knot

Testimony #38

"He hit on me repeatedly, which freaked me out b/c he was much older and had all the power."

Final result

Testimony #38

Grid legend

1: At conference, No. 2: Hierarchy, Lower. 3: Hierarchy, Mixed. 4: Stayed in academia, Yes. 5: Incident reported, No. 6: Incident reported, Yes. 7: At conference, Yes.
8: Report successful, Yes. 9: Report successful, No. 10: Avoided people or places, Yes. 11: Avoided people or places, No. 12: Stayed in academia, No. 13: Type of abuse, Both.
14: Type of abuse, Verbal. 15: Type of abuse, Physical. 16: During business hours, No. 17: During business hours, Yes. 18: Hierarchy, Higher. 19: Hierarchy, Equal.

Final result

He hit on me repeatedly, which freaked me out b/c he was much older and had all the power.

Artificially added control points to improve stability in curve extremities
VisAP’20, Pictorials and annotated portfolios.

Present very similar data values and for this reason scattering is done with a force layout that starts with assigned coordinate positions and uses collision detection to restore the visibility of overlapping elements. This solution to position all reports in the Cartesian plane, allowing for a free and open-ended exploration (the audience can zoom and pan within the space). In addition, informal interviews with friends, colleagues and students confirmed that the visualization strategies adopted are capable of raising curiosity and inviting the audience to dedicate time and reflect.

However, the visualization was still incomplete, because at this point only stories excerpts were displayed. To cope with the need of accessing the complete data, we inserted the ability to ‘unravel’ knots, look at the data-values that generate a particular shape and read the complete report. The information about victims’ and perpetrators’ status were left as annotations that appear in the corresponding modes.

**Modes and user interface**

To avoid information overload and to allow for a more effective self-identification, we spread the available data onto three modes of exploration that correspond to the three
positioning possibilities previously described. The first and main mode is recognizable from the yellow background and corresponds to the positioning of knots that is based on incident descriptions. In this way knots with a similar tagging in terms of incident description (and therefore with similar shapes) are positioned close to each other. The second mode positions knots based on victims’ status and it is recognizable from the white background. The third one uses the harassers’ status as positioning criteria and has a dark background. The user interface has a switch (positioned in the bottom center of the screen) to allow the audience to switch between the three exploration modes. It also presents a zoom control (on the left) and a filter designed on top of the control-points grid (bottom-right corner) that allows users to locate knots with specific data values.

The visual language of Tied in Knots is designed to recall an archive, in order to highlight the curatorial components of our project. Archives, entail curation, especially in the form of judgement and selection. For this reason we embedded testimonials onto cards and we chose a monospaced typeface (Inconsolata) to evoke a connection to typewriters and library card systems. This font also renders well in the knotted texts and in their unravelling. The color choices follow a different logic. We chose black text on white background for victims because they really put memories down in black and white by participating in the survey and we used white on black for harassers to stress out their diametrically opposed position. Yellow is the accent color.
we chose for incidents, because it is a tone that indicates the need for attention and because it matches well with the rest of the color palette.

**Voices**
The last feature added was audio recordings, created thanks to volunteers that read story excerpts aloud. Its aim is two-fold. On the one hand, human voices are able to emphasize the emotional impact of the visualization and, in the case of an installation in a public space, they can also draw attention and raise curiosity. On the other hand, the fact that people dedicate individual time to read, digest, and record, represents an important and innovative way to show support and manifest personal concerns for the problem of SH in the academy. After careful consideration of the appropriateness of having men participate in reading aloud stories most often submitted by women, we decided to invite anyone willing (all genders) to participate as a form of solidarity. We believe that similar actions can result in a feeling of empowerment for victims. Currently, the majority of volunteers have been males and this might sound disorienting, but we are considering the possibility to further crowdsourcing more voices.

**Discussion**
In this project we leveraged approaches of different disciplines with the aim of using data visualization to communicate and promote
Acknowledgements
We would like to acknowledge all the people who participated in the survey and recalled distressing experiences in the process of sharing their stories. Furthermore we want to thank colleagues from Vialab, Ontario Tech and from DensityDesign for the help and their suggestions.

References

1. K. Kelsky, “A Crowdsourced Survey of Sexual Harassment in the Academy,” The Professor is In, Apr. 12, 2017. https://theprofessorisin.com/2017/12/01/a-crowdsourced-survey-of-sexual-harassment-in-the-academy/ (accessed Jun. 04, 2020).
2. C. D'Ignazio and L. F. Klein, "Feminist Data Visualization," 2016.
3. M. Correll, "Ethical Dimensions of Visualization Research," in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI ’19, Glasgow, Scotland Uk, 2019, pp. 1–13, doi: 10.1145/3290605.3300418.
4. M. Doric, P. Feng, C. Collina, and S. Carpendale, “Critical InfoVis: exploring the politics of visualization,” in CHI ’13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA ’13, Paris, France, 2013, p. 2189, doi: 10.1145/2468356.2468739.
5. J. Drucker, “Humanities Approaches to Graphical Display,” Digit. Humanit. Q., vol. 5, no. 1, pp. 1–23, 2011.
6. J. Boy, A. V. Pandey, J. Emerson, M. Satterthwaite, O. Nor, and E. Bertini, “Showing People Behind Data: Does Anthropomorphizing Visualizations Elicit More Empathy for Human Rights Data?,” in Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver Colorado USA, May 2017, pp. 5462–5474, doi: 10.1145/3025453.3025512.
7. M. Mauri, "Progettare il non-finito," 2015.
8. M. Mauri, T. Elli, G. Caviglia, G. Uboldi, and M. Arzi, “RAWGraphs: A Visualisation Platform to Create Open Outputs,” Proc. 12th Bıaınn. Conf. Ital. SIGCHI Chapter, p. 28:1-28:5, 2017, doi: 10.1145/3125571.3125585.
9. P. Cruz, J. Wibbey, A. Ghael, F. Shibuya, and S. Costa, "Dendrochronology of U.S. immigration," Inf. Des. J., vol. 25, pp. 6–20, Dec. 2019, doi: 10.1075/idj.25.1.01cru.
10. A. Malka, “HE LIKED THICK WORD SOUP. Reading Ulysses with your fingers,” http://chronotext.org/, 2014. http://chronotext.org/Ulysses/index.htm (accessed Aug. 05, 2020).
11. C. Frayling, "Research in art and design," Res. Pap., vol. 1, no. 4, p. 9, 1993.