Cartographers of North Korea

Wonyoung So*

Department of Urban Studies and Planning, Massachusetts Institute of Technology, MA 02139

ABSTRACT

The focus of this project is to discuss the collaborative mapping strategies of contributors in uncharted territories, using North Korea as a case study. OpenStreetMap (OSM) enables "armchair mappers" to map territories such as North Korea in which the government has controlled internet access to its residents. Here, I ask the questions of who is mapping to North Korea, which tools and methods the contributors use to have access and represent North Korea, and which are the motivations behind such mapping endeavor. To do this, I analyze the technical aspects of OSM data in North Korea, and analyze the structured correspondence I exchanged with 889 contributors.

Keywords: Cartography, North Korea, Text Visualization, OpenStreetMap, Volunteered Geographic Information

1 INTRODUCTION

Mapping is a combination of technical, political, and social capabilities and intentions. It has been so throughout the history of cartography. Technical aspects encompass the instruments to see, measure and represent spatial features, as well as the scientific knowledge of the time, usually translated into cartographic standards. The political aspects of mapping connote the intentionality underlying the creation and use of maps, whereas the social aspects include the ways and features the mapping and mapped groups living in certain areas see space, sometimes eliding or pushing the boundaries of established cartographic standards.

The focus of this paper is to discuss the collaborative mapping strategies used by open-source mapping contributors, using North Korea as a case study. Here I ask the basic questions of who is mapping North Korean cities, which tools and methods the contributors use to have access and represent these cities, and which are the motivations behind such mapping endeavor. In order to do this, I analyze the technical aspects of OSM in North Korea, and analyze the structured correspondence I exchanged with 889 contributors.

The visualization, which is produced for understand and display the strategies, aims to tackle the challenges related with digesting the information that is in the correspondence I exchanged with 889 contributors. Text data is generally considered to be hard to be visualized, especially without any contextual information. To solve the problem, I give the users contextual information— showing interesting Points of Interests (POIs) and descriptive statistics about the OSM data in North Korea. Then the visualization shows the correspondences, by reducing the dimension of text and categorizing every sentence.

The paper continues by describing the efforts to mapping North Korea's territory and unveiling ordinary aspects of urban daily life after the country has become overly closed to the outside world. It follows with a discussion of contemporary mapping tools, in particular OpenStreetMap, and then it presents the methodology employed in this research, the data gathering and analysis. Departing from an apparent technical feature of contemporary mapping, such as online mapping, the paper discusses how the use of open source mapping tools, in a particularly secluded territory, involves technical, political, and social aspects inherent to mapping.

* wso@mit.edu

2 RELATED WORK

To visualize continuous contribution of spatial and temporal data, a number of attempts have been done in the field of data visualization. In particular, there are numerous data visualizations about OpenStreetMap's history of change [1, 2, 3]. Since it is spatial and temporal data, which involves the date of change, it is mostly generated as an interactive data visualization or a video. Compared to former visualizations which are lack of showing individual changes, an attempt has been made to show individual OSM contributors in Seoul. [4]

On the other hand, there have been some attempts to visualize text data. However, compared to quantitative data visualization where research has been conducted on what kind of chart to use in what situations and these efforts have been published by media outlets such as Financial Times [5], studying about text data is still in preliminary state and there has been no meaningful visual vocabulary for visualizing text data. It ranges from counting words [6], producing a tag cloud [7], and calculating tf-idf to figure out uniqueness of such words [8]. However, although it could help gisting of what is going on certain document, all of the method has one same problem; it shows only one or two words and detach each word from the sentence or paragraph which originally contains the words. Users would lose numerous contexts and meanings from those processes.

Automatically applying visual styles haphazardly also could be problematic. In the field of Natural Language Processing (NLP), automatically analyzing text and annotating them is such an important topic because they have dealt with a large amount of text data. Strobelt et al, tested what is the most effective visual style when annotating such text data [9]. However, it is problematic to apply to practical data visualization because if various visual styles applied to text data, the text would be hard to be read. In particular, with the respect to the fact that data visualization is multi-disciplinary producing activities where graphic design—especially typography in this context—should be heavily involved, there is a clear problem with varying the text style just for distinction in the visualization of data without sufficient thought.

However, in spite of those difficulties, there are some notable text data visualization projects done by data visualization practitioners and journalists. The Preservation of Favoured Traces, done by Fathom Information Design, visualizes changes of Charles Darwin's the Origin of Species [10]. The book had been revised six times, and it displays which version is changes which part very clearly because this project reduced text dimensions into six editions and encoded each edition as distinct colors. Although it is impossible to read the content inside the visualization, it conveys the intended information concisely. One of the other notable attempts is visualizing redacted area of confidential documents. After the Muller report was released, a number of journalists tried to visualize what content is inside of the report. However, one of the most effective visualization methods was, unexpectedly and ironically, just showing the redacted area of the report [11]. This case also cannot read any content inside the report, but it clearly conveys the problem of how the report is complete and shows what how much area is hidden.

To gist text data, reducing dimension of text data appears to be indispensable. However, in the case of visualizing correspondence, it is also crucial to read each letter and make a sense of their motivation of why they mapped North Korea. Thus, using an interactive interface of

visualizing the correspondence by categorizing each sentence, and showing the area of each contribution, I aim to increase the sense of what each contributor intends to do.

3 METHODS

The methodology is divided in two main parts: the descriptive statistics of the contributors, and interviews with contributors. The OSM data of North Korea was dumped at October 2018 using Geofabrik's OpenStreetMap Data Extracts (http://download.geofabrik.de/), a service that aims to split the OSM Planet data to a country-level and update it daily.

In the descriptive statistics of the contributors, I analyze the number of contributions per contributor, the average number of contributions, the country of origin of the contributor. Additionally, I analyze the types of contributed data points, such as street-level POIs (restaurants, statues, parks), and controversial data points (such as nuclear test sites). Finally, I analyze the changesets of each contributor.

In the second part, I sent a letter via OpenStreetMap's messaging system to all 889 contributors, and follow-up interviews to all contributors who replied the initial letter, with questions about, but not limited to, their motivation, nationality, and opinions of North Korea and mapping strategies. I sent the message to all 889 contributors, and got 222 responses.

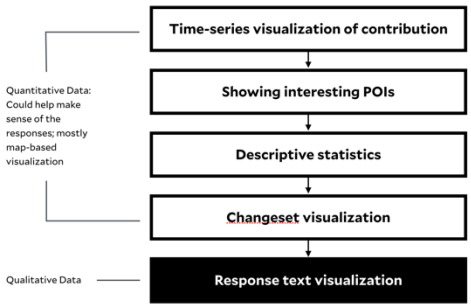

Figure 1: The story structure: it starts from showing quantitative data and end with displaying qualitative data.

The visualization, a result of the research, utilized scrolling interaction to follow linear storytelling structure. However, with some guidance from the text box, users can interact with the visualization to find in-detail information. The visualization shows, first, quantitative data by showing data on the map and graph, then with the context, it shows the correspondence with contributors. shows time-series visualization of contribution and displays interesting POIs contributed in North Korea, such as military camps or nuclear test sites. After it displays the changeset data, it shows the response text data, which is categorized of It was produced using Mapbox GL JS, a mapping framework, and D3.js in the React JS framework.

4 RESULT

4.1 Descriptive statistics of contributors: Who are the contributors and how did they contribute?

As Figure 2 shows, the total number of contributors as of 2018 was 889, and they have contributed 324,415 data points since 2008. The figure 3 shows that the distribution roughly follows a power-law distribution, and the top 10 contributors drew 61 percent of the map.

Since OSM does not include each contributor's demographics such as country of origin, gender, or age, I could estimate the contributors' characteristics using relevant data that OSM has. One possible estimation is how many contributors use Korean; it is interesting to note that of the contributors who labeled the map objects, 62 percent labeled them in Korean.

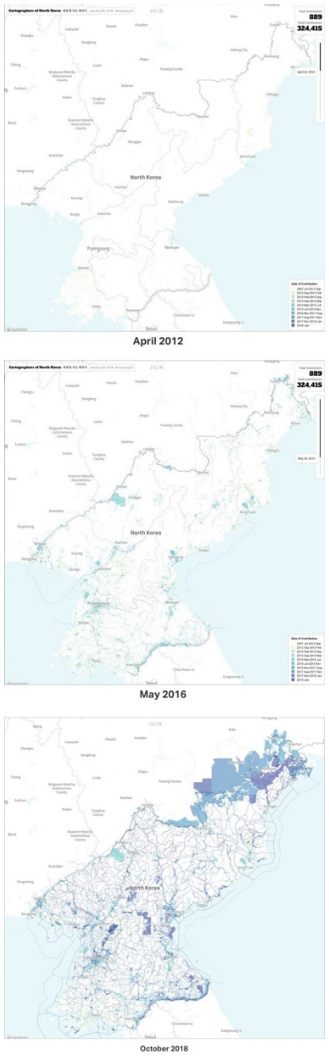

Figure 2: Contribution changes of the OSM data in North Korea from April 2012 to October 2018.

One of the other ways of estimating their activity is through OSM changesets, which is a history of past contributions of each user. Using the changesets, it can be seen which regions, other than North Korea, have been the contributors' interests. When summing up every changeset of 887 contributors, it does not give much insights, although there are some countries that have been in good relationship with North Korea such as Germany, China, and Russia. However, there are some countries coming up with heavy contributors. For instance, the top 20 contributors who use Korean also contributed to the countries as follows: China, Germany, India, United States, and South Korea. If the number of contributors is reduced to 5, the list of countries changes as follows: India, Germany, Ukraine, Russia, and Japan.

It is particularly interesting to note that there are a number of street-level Point of Interests (POIs), including one of the leader's graves ("The graves of Mr. Kim Hyeong-jik and Mrs. Kang Ban-seok"), restaurants with specific cuisines (Italian, Chinese, rice-cake soup, and so on), Kim Il Sung's Memorial Statue, and so on. Typically, this kind of local information can only be accumulated only by local residents. Moreover, there are a number of attempts to reveal secrets of the North Korean regime including nuclear test sites, military sites, even the location of individual fighter planes.

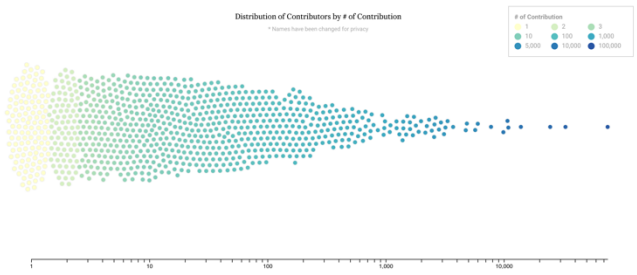

Figure 3: Distribution of contributors by the number of contributions. It shows that the top 10 contributors drew 61 percent of the map.

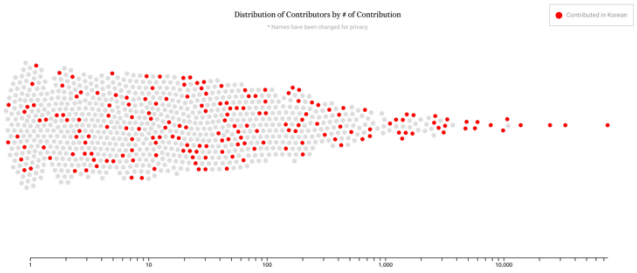

Figure 4: Contributors who labeled data in Korean, which is 62 percent. In particular, all of the top 5 contributors labeled data in Korean.

4.2 Interviews

For visualizing interviews, I initially grouped the answers into three categories: motivation for mapping North Korea, particular interest in North Korea, and about mapping strategies. After a brief examination, motivations for mapping North Korea turned out to be vary from mapping as a private hobby to a political act. Then, all of the responses were split into sentences and labeled into 13 different categories. For "Why?" section, it comprised of 9 reasons: It's for fun, After exposed from media, Humanitarian Purpose, Have Personal Attachment, Search for empty space, Mysterious (thus interesting) Place, Advocating Open Mapping, Fixing things around the world, and Learn about foreign country. For "How?" section, it comprised of 3 ways: Satellite Imagery, They've been to there, From the Internet. For comments about reunification of Koreas or general notes about North Korea, it also has categories for each. In the text visualization part, it renders all correspondence in one screen, to be able to check immediately how much the space is occupied according to the selection the users made. If the users would like to know more about individual responses, then they can click each response to see not just each text, but individual contribution histories.

These kinds of combining linear sequence and exploration (so called Martini Glass Structure) benefits users because they already went through all of the stories and contexts that is needed for understanding what each category means, and even in the middle of exploration, the visualization provides examples of the exploration method, such as highlighting the different method of mapping (satellite imagery, they have been to there, and combining with resources in the internet). Therefore, the users fully understand and benefit from the rich given information and finally can explore more deeply and actively engage by reading individual responses.

5 Conclusion and Future Work

I examined the OSM data of North Korea by analyzing the OSM data in North Korea and interviewing with over 200 contributors, and identify their motivation and concern when the mapped North Korea. The cartographers of North Korea are mostly outsiders from North Korea. However, they are resonated with distant issues and act locally. They have empathy over distance. They combined and entangled information from different sources like investigative journalists, thus producing the place-

Figure 5: Visualization of Responses: According to the selection user makes, it highlights sentences that matched with selected categories in the paragraphs.

level knowledge. By contributing to the remote territory, they obtain local knowledge. We need to rethink about how to accumulate local knowledge, and how to generate reaction over distance and derive active engagement in crowd-sourced projects. We should also think about who finally benefits from all of the process, since the mapping is a combination of technical, political, and social capabilities and intentions, thus often resulting in unexpected outcomes.

For a visualization perspective, Visualizing text data is such a challenging task, and the current method applied to this paper is limited because it requires time-consuming labor. However, regarding the current limited text visualization techniques and inherent limitations of text data, detaching words from sentences or paragraphs is not a feasible solution as it loses context. Thus, if a data visualization designer wanted to apply current text visualization methods, it would be super important for them to think about the tradeoffs where the more they want to detach, the more they lose in terms of content. Thus, I argue that reducing dimensions of text data and visualizing that data with context would be a good solutions of visualizing text data. And if a machine-learning model could extract that categorization from the text, thus automating the categorization process, that would be helpful.

References

[1] "10 years of OpenStreetMap," *10 years of OpenStreetMap*. [Online]. Available: https://mapbox.com/ten-years-openstreetmap/. [Accessed: 12-May-2019].

[2] ItoWorld, *US edits to OpenStreetMap 2007-2012*. 2012.

[3] Rethans D, 2014, "OpenStreetMap: A Year of Edits (2013)", https://vimeo.com/83164362

[4] So W, “2017 Seoul OSM Dashboard: Seoul Libre Maps”, https://wonyoung.so/2017-seoul-osm-dashboard
[5] Financial Times, “Visual Vocabulary”, http://ft-interactive.github.io/visual-vocabulary/
[6] Wordcount · Tracking the Way We Use Language, http://www.wordcount.org/
[7] Davies J, “Word Cloud Generator”, https://www.jasondavies.com/wordcloud/
[8] S. Ahern, M. Naaman, R. Nair, and J. H.-I. Yang, “World Explorer: Visualizing Aggregate Data from Unstructured Text in Geo-referenced Collections,” in *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries*, New York, NY, USA, 2007, pp. 1–10.
[9] H. Strobelt, D. Oelke, B. C. Kwon, T. Schreck, and H. Pfister, “Guidelines for Effective Usage of Text Highlighting Techniques,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 489–498, Jan. 2016.
[10] “Fathom | The Preservation of Favoured Traces,” *Fathom Information Design*. [Online]. Available: https://fathom.info/traces. [Accessed: 12-May-2019].
[11] J. Schleuss, “See how much of the Mueller report was redacted,” *www.latimes.com*. [Online]. Available: https://www.latimes.com/projects/la-na-mueller-investigation-report-trump-redaction/. [Accessed: 12-May-2019].