Geographies of Meaning: Visualizing Russian Émigrés and Travelers

by

David Kurkovskiy

A thesis submitted in partial fulfillment for the degree of Bachelor of Arts

in the Department of Computer Science

Yale College

December 2016
This project theorizes and experiments with digital humanities practices for visualizing geographic movement of six Russian poets from 1914-1941. The project explores theoretical writings about the place of digital methods in humanities departments, as well as how technical tools allow humanities researchers to reimagine both old and new forms of knowledge production. Two main visualizations of the target poets are produced at the conclusion of the project: a multimedia view juxtaposing a map with representations of cultural production for each poet and a storytelling view narrating the biographies and physical journeys of individual poets. Finally, the paper concludes with a specification for possible methods of including semantic information (in the form of textual analysis) as part of a map-based visualization. While the initial two experiments demonstrate the usefulness of visualization in envisioning old genres, the specification theorizes a new possibility for knowledge production unique to the medium of map-based visualization.
Acknowledgements

I would like to sincerely thank Professor Holly Rushmeier for all the support she offered throughout this semester, not only for her advisement on the project and on my visualization experiments, but for always encouraging my explorations at the intersection of computer science and the humanities. Our conversations on the opportunities and challenges of the digital humanities inspired this project and my ambition to keep exploring. I would also like to thank Professor Dana Angluin, my first Computer Science instructor at Yale, for always lending an ear and encouraging me to double major in two subjects that could not be farther apart (or so we thought). I am also grateful for my family in the Slavic Department, for always reminding me the value of interdisciplinary and inquiry. Especially Marijeta Božović for teaching the DH Lab and advising the humanities portion of our project. And my deepest gratitude to my teammates (Emalin Kelso, Stephan Svishnikov, and Aleksandra Konarzewska) in the DH lab for their support in our countless hours of hacking, scraping, curating, and researching chronicles of our beloved poets.
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>i</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>ii</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1</td>
<td>Digital Humanities and Maps</td>
</tr>
<tr>
<td>1.1.1</td>
<td>Social Networks</td>
</tr>
<tr>
<td>1.1.2</td>
<td>Visualizing Geographies</td>
</tr>
<tr>
<td>1.2</td>
<td>Visualizing Soviet Émigrés and Travelers</td>
</tr>
<tr>
<td>1.2.1</td>
<td>The Problem</td>
</tr>
<tr>
<td>1.2.2</td>
<td>Tools and Limits</td>
</tr>
<tr>
<td>1.2.3</td>
<td>Conclusions</td>
</tr>
<tr>
<td>2 Literature Review</td>
<td>5</td>
</tr>
<tr>
<td>2.1</td>
<td>What are the Digital Humanities?</td>
</tr>
<tr>
<td>2.2</td>
<td>Close Reading and Distant Reading</td>
</tr>
<tr>
<td>2.2.1</td>
<td>Machine-assisted Reading</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Reading at a Distance</td>
</tr>
<tr>
<td>2.3</td>
<td>Media and Visualization</td>
</tr>
<tr>
<td>2.4</td>
<td>Geographies in Context</td>
</tr>
<tr>
<td>3 Tools and Methodology</td>
<td>9</td>
</tr>
<tr>
<td>3.1</td>
<td>Datamaps</td>
</tr>
<tr>
<td>3.2</td>
<td>Data Protocols and Procedure</td>
</tr>
<tr>
<td>4 Experiments and Results</td>
<td>11</td>
</tr>
<tr>
<td>4.1</td>
<td>Multimedia View</td>
</tr>
<tr>
<td>4.2</td>
<td>Storytelling View</td>
</tr>
<tr>
<td>4.3</td>
<td>Designing a Semantic View</td>
</tr>
<tr>
<td>5 Conclusion and Next Steps</td>
<td>15</td>
</tr>
<tr>
<td>5.1</td>
<td>Conclusion</td>
</tr>
<tr>
<td>5.2</td>
<td>Next Steps: the Afterlife of Visualization</td>
</tr>
<tr>
<td>Bibliography</td>
<td>17</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

The application of digital methods to humanities scholarship aims to engender the work of humanists, augmenting traditional forms of humanities and social science analysis in addition to carving out the possibility for new forms of analysis. For example, topic modeling a corpus of literary texts might help the humanities scholar find confirmation for patterns, the existence of which could be hinted by traditional close reading methods. Alternatively, such a statistical modeling could allow the humanists to discover new patterns that were not previously identified by traditional scholarship methods.

The potential contributions of technology to humanities inquiry are thus twofold: on the one hand, technical models and visualizations facilitate traditional humanities methodologies by either streamlining the process or juxtaposing different kinds of available information to allow the humanist to come to the same conclusions about a fixed set of subjects in a faster or more efficient way; on the other hand, technological methods can shed light on dimensions of the humanist’s inquiry that would not otherwise be discoverable by traditional humanities practice. The humanist (and the technologist) in turn must be equipped with the proper training to analyze these results (especially in the latter case) and apply them to make conclusions accessible to the broader humanities community. Digital humanities technology is not a replacement for the humanist; it can serve as a prosthetic for cultural and knowledge production that still requires the input and interpretation of the scholar. Moreover, digital humanities tools—when created with the consideration of more traditional humanists and humanities inquiry—can lead to unique forms of knowledge production and cultural capital. Thus they not only allow humanists to do their work better, but they offer a potential for the expansion of attainable cultural knowledge, revolutionizing traditional understandings of cultural capital.
1.1 Digital Humanities and Maps

1.1.1 Social Networks

Maps and networks have become a particularly fruitful medium of experimentation at the intersection of humanities and technological disciplines in recent years. Social networks, for instance, have been the subject of many studies in both computer science and humanities. They have also become a tool for humanists to reimagine their scholarship in new ways. To take a simple example, many literature instructors have been to envisage the interrelationship of characters in a novel as nodes on a social network graph, encouraging students to imagine (and create) personal Facebook pages for individual characters; these might involve depicting which characters were acquainted with one another, what they wrote, how their relationships with one character differed from relationships with another. While such an example demonstrates a simple use of a digital tool in humanities practice, it demonstrates the ability of such tools to visualize useful knowledge traditionally targeted by humanities scholars with a novel, digital interface. Such a visualization might help the student or scholar identify patterns of communication between characters (which characters in a book did not engage in a single conversation?).

1.1.2 Visualizing Geographies

Other recent projects highlight more than just interrelationships along a graph, but focus more on depicting changing realities over time and geographic location. For example, the Harvard Imperiia project directed by Department of History Professor Kelly O’Neill uses the GIS mapping tool to study the Russian Empire’s spatial history, making a range of different maps available and promoting the use of historical maps as a pedagogical tool in the history classroom. The project attempts to not only facilitate orthodox cartography-based practices through technical meals (e.g. calculating the area of a province), but to make possible new forms of geographic inquiry that were not previously possible (use statistical methods to calculate distances between villages and towns throughout different provinces). So as in the above example of topic modeling for the digital humanists, map-based visualizations can serve to either facilitate the discovery previously researchable information or uncover otherwise undetectable patterns. Visualizing maps can thus help humanists more efficiently produce previously attainable knowledge and make available knowledge that would not previously be accessible.
1.2 Visualizing Soviet Émigrés and Travelers

1.2.1 The Problem

This essay is an exploration of the potential of maps and geography-based visualization as a digital humanities practice, focusing specifically on Russian travelers and émigré authors in the late imperial period and early Soviet period. Authors were chosen because of the potential of juxtaposing artistic production, interrelationships between authors, and the qualities of their written work and displaying one or all such aspects on an interactive map. The project achieves two main goals: first, it involves the implementation of several visualization experiments depicting the travels, biographies, and cultural production of six Russian émigrés and travelers (Anna Akhmatova, Osip Mandelshtam, Vladimir Mayakovsky, Vladimir Nabokov, Boris Pasternak, Marina Tsvetaeva) from the years 1914-1941. Second, it theorizes the landscape of possibilities for digital humanities projects in the study of transnationalism, demographic movements, intellectual history, and the study of the early Russian émigré movement during the early Soviet period.

1.2.2 Tools and Limits

While the technical experiments designed in this project serve as positive examples of possibilities for visualization that could serve as pedagogical tools for instructors or as technical supplements for humanities researchers, the conceptual portion of the project explores both the possible contributions and the theoretical limits of digital humanities as a mode of inquiry in the humanities. The project considers the possibilities of map-based visualizations in terms of the initial overview of possible digital humanities projects detailed at the beginning of the essay: possibilities to describe previously observable patterns, and to reveal unobservable patterns in the input information to produce new forms of humanities knowledge.

1.2.3 Conclusions

The essay concludes that map-based geographic visualizations can help visualize knowledge and cultural production in ways useful for humanities scholars. The two initial experiments implemented for the project—the multimedia and storytelling views—demonstrate the applicability of a simple map visualization software for creating digital humanities tools that mimic the genre of museum exhibition and biography. But the conceptual contribution of this essay is points to the challenges of map-based visualization in achieving
the second kind of digital humanities possibility: uncovering the (previously) unknowable. In order to provide geographic information that is previously unknown to the scholar, the map must correlate changing geographies with semantic significance. This semantic analysis could take different forms: pictorial analysis of contemporaneous pictures or sound analysis of contemporaneous recordings or statistical analysis of involved texts. Though a full implementation of such a semantic visualization was not completed at the time of writing, the importance of such a methodology for digital humanities practices (especially for the latter possibility of digital humanities practice) is expanded upon in the paper as a conclusion that is equal in importance to the currently implemented visualization prototypes.
Chapter 2

Literature Review

The following literature review features a brief overview of digital humanities practices as theory, with special attention to the treatment of networks and graphs in digital humanities projects. In so doing, we provide a definition for “digital humanities” that theorizes its techniques as a multidisciplinary practice useful for humanists and computer scientists alike. The goal of this review is to show the potential of such tools in humanities inquiry, in addition to suggest forms of humanism that have always been a part of human-technology interactions.

2.1 What are the Digital Humanities?

Matthew Kirschenbaum describes “digital humanities” as a movement or methodological outlook, rather than one set of texts or technologies. But in addition to a sum total of technologies and technology-based methods in university humanities departments, Kirschenbaum points to digital humanities as a form of knowledge production and social network in itself. He writes, “Digital humanities is also a social undertaking. It harbors networks of people who have been working together, sharing research, arguing, competing, and collaborating for many years.” He points to opportunities for digital humanities initiatives for attracting research funding and creating new forms of humanities inquiry.[1]

Joanna Drucker theorizes the possibility of creating a “humanistically informed theory of the making of technology (a humanistic computing at the level of design, modeling of information architecture, data types, interface, and protocols.) This emphasis on what Drucker calls “humanistic” program design reminds us that digital humanities methodology cannot be reduced the use of pre-packaged software by humanities scholar, but
should involve a direct and academically mutual relationship between the humanist and the technologist. For this reason, the current essay so strongly emphasizes a focus on technical design that is sensitive to the need of humanists, and relies on constant feedback from humanists at every step of design (developing the data model, prototyping the visualization, etc.). [2]

2.2 Close Reading and Distant Reading

2.2.1 Machine-assisted Reading

Katherine Hayles also makes an important contribution in theorizing digital humanities as a multidisciplinary methodology. She explains the potential of digital tools to assist students with humanistic reading by pointing to the educational principle codified by L. S. Vygotsky: the “zone of proximal development”, which describes the “distance between instruction and available skills must be capable of being bridged, either through direct instruction or, as Vygotsky notes, through working with “more capable” peers. Though Vygotsky’s theory specifically refers to the pedagogical benefit of working with peers or instructors, Hayles argues the same concept is at play with digital tools that supplement traditional close reading strategies with machine reading. She argues that machine- and hyper- reading strategies are just as necessary to literary studies as traditional close reading. She envisions a new form of literary studies that teaches “literacies across a range of media forms, including print and digital, and focuses on interpretation and analysis of patterns, meaning, and context through close, hyper-, and machine reading practices.” [3]

2.2.2 Reading at a Distance

Franco Moretti, a Stanford Professor famous for laying the groundwork for much of digital humanities methodological theory, described the zoomed-out, machine-assisted method of reading large corpora of texts as “distant reading.” He writes, “Distant reading: where distance, let me repeat it, is a condition of knowledge: it allows you to focus on units that are much smaller or much larger than the text: devices, themes, tropes—or genres or systems. And if, between the very small and the very large, the text itself disappears, well, it is one of those cases when we can justifiably say, Less is more. If we want to understand the system in its entirety, we must accept losing something.” Moretti’s formulation of machine-assisted distant reading is directly relevant for the task of visualizing and depicting artist travels on an interactive map. The scarcity of physical space on a computer screen limits, for instance, the amount of information that can be
depicted in a single visualization. Furthermore, superfluous or noisy data risks colluding the visualization with information that will prevent the humanist from conducting his/her inquiry rather than augmenting the research. But viewing the data at a distance offers the benefit of noticing new patterns and correlations that would not previously come to the surface. Distant reading, as Moretti later suggests, also provides an opportunity to reject the bias of particular canonical texts and historical events as disproportionally representative. Consider the example of Soviet poet Vladimir Mayakovsky, whose travels are depicted in this project. A great amount of literary work on the poet has highlighted his visits and writings of America, necessarily focusing less on his other travels. The benefit of a “distant reading” as in the implemented visualization is an opportunity to consider other moments of history and literary production with equal weight. By weighing all years equally and displaying representative works and historical narratives throughout the period of inquiry (1914-1941), the project offers a distant reading of stories that are usually read “closely” in the scholarship. [4]

2.3 Media and Visualization

Bolter and Grusin write about the importance of interactive graphic user interfaces for humanities inquiries grounded in media studies theory. They define and theorize three interrelated concepts of human-computer media interaction: immediacy, hypermediacy, and remediation. Immediacy defines the contact point between the medium and what it represents (e.g. the still of a visualization that represents individual nodes), while hypermediacy involves the creation of a multimedia space. According to Bolter and Grusin, “In every manifestation, hypermediacy makes us aware of the medium of media and (in sometimes subtle and sometimes obvious ways) reminds us of our desire for immediacy. This formulation is directly applicable to the multimedia view of the current project. Though the initial representation of the geographic location of each poet in a given year demonstrates an initial connection between our visualization medium and its object of representation (“immediacy” by Bolter and Grusin’s definition), the availability of video clips, voice recordings, and visual images that become available once hovering over individual nodes of the map create a multimedia experience that disrupts the simple “immediacy” of the view at first site (this, arguably, enacts the phenomenon of “hypermediacy”). Furthermore, the multimedia view implements Bolter and Grusin’s third concept of “remediation”, defined as the “repurposing” of one medium form by another. They write, “With reuse comes a necessary redefinition, but there may be no conscious interplay between media. The interplay happens, if it all, only for the reader or viewer who happens to know both versions and can compare them.” The last definition speaks especially to the importance of the observer (user) of the map interface. It is through
the user’s interaction with the nodes and the different media they unfold into that the
typical experience of immediacy is transformed into one of hypermediacy and remedita-
tion of the original source materials. Though the theoretical applicability of Bolter and
Grusin’s is not strictly limited to graphic interfaces, they help chart out a foundational
framework for understanding the importance of juxtaposing different kinds of media in
a graphic interface. [5]

2.4 Geographies in Context

Finally, Scott Weingart writes about the importance of contextualizing networks with
geographic information in graphical visualizations use by digital humanities scholars.
He writes, “By combining geography, point statistics, and networks, we can create base
maps against which we can contextualize whatever we happen to be studying. This is
just one possible combination; base maps can be created from any of a myriad of sources
of data. The important thing is that we, as humanists, ought to be able to contextualize
our data in the same way that we always have.” He adds that even if these maps do not
produce knowledge that was not previously available, they “are still useful for at-a-glance
synthesis of large scales of information, or learning something new about areas one is not
yet an expert in.” With proper contextualization, however, a map-based visualization can
do a lot more than assist humanists in synthesizing information. With the introduction
of additional information, they can allow scholars to study the interrelationship between
geographic location and other phenomena (such as the amount of letters written, the
prevalence of certain words, themes, or sentiments in texts). [6]
Chapter 3

Tools and Methodology

The visualization experiments were generated with the help of the Datamaps library powered by d3.js. Javascript and JQuery were used to add event listeners such as buttons to advance and recede a year in the multimedia view.

3.1 Datamaps

Datamaps is a library that provides pre-generated d3 topojson maps on which information in the form of SVG circles can be placed (“bubbles”). The library also allows custom topojson maps to be used. The library was chosen because of the ease of charting nodes with known latitude and longitude coordinates on the map. On-hover events were used to attach additional information to the map nodes, such as short clips, voice recordings, and pictures in the multimedia view. This allowed the addition of an extra dimension to the map, in addition to time and geographic coordinates.

3.2 Data Protocols and Procedure

In collaboration with a group for the Fall 2016 Digital Humanities: Avant Gardes and Émigrés lab, CSV sheets were created for integration with the map visualization. The two sets of CSV sheets that were drafted thus far include: a multimedia sheet that features longitude/latitude coordinates, location name, and a relative link to a related photo (*.jpg), video clip (*.mp4), or voice recording (*.mp3) for each author, year tuple of interest; and separate narrative sheets for each author that features longitude/latitude coordinates, location name, the order of the trip (relative to the current year for the current author), and a short narrative description of the author’s activity during the
trip. The latter CSV sheet allowed for multiple trips of a single author in a given year to be encoded, and does not necessarily feature an entry for each author in every year. The former sheet, however, included a location for every valid year, author tuple. The CSV sheets were accessed directly by the d3.csv command, excluding the need for alternative database storage and a backend site for the current project. It is also worth noting that the compiled data would not have been available were not for the efforts of the group in the DH Lab course.
Chapter 4

Experiments and Results

Two full map visualization views were implemented as part of the project. The full code was pushed to GitHub, and the public version of the site was published using GitHub pages. Though the third view incorporating semantic analysis of the poets’ texts was not completed at the time of writing, the final section of this chapter features proofs and a technical design specification for a map view incorporating either topic models of contemporaneous texts and/or textual analysis of word frequencies.

4.1 Multimedia View

![Multimedia View](image)

**Figure 4.1:** Multimedia view mapping artists with images, clips, and recordings

The first experiment implemented was the multimedia view inspired by Bolter and Grusin’s aforementioned theorization of hypermediacy and remediation in graphical interfaces. The view features a node for each poet, year tuple, depicting the traveler’s
geographic location (a single location is chosen in this view per year). Most of the included nodes include a media sample that in some way relates to the author’s activities in the indicated location and/or year: video clips, sound recordings, and photographic images. This view can be understood as a digital interpretation of a curated museum exhibit, allowing technologists and humanists to work together to choose representative works and media samples.

Use: hover over the nodes to see the attached media. The bar at the bottom of the page can be used to cycle between the years that were studied (1914-1941). As shortcuts, pressing the 'M' key proceeds a year and the 'Z' key moves back a year.

4.2 Storytelling View

The second experiment implemented was the storytelling view. Given that the first view forces the technologist and humanist alike to choose a single set of geographic coordinates for every year in the inquiry, this view was created to introduce more information into the map visualization without inundating users with too many nodes and risking illegibility of the project. The storytelling view also involves narratives that accompany each "scene" (if poets traveled to multiple locations throughout the course of a year, multiple scenes (numbered below the year at the top of the page) appear. This view can be seen as a digital instantiation of a biography or chronicle genre, allowing scholars to envision episodic histories of each poet in order with the aid of a map representation.

Use: click the name of an author and click the spacebar to proceed to the next scene.
4.3 Designing a Semantic View

Finally, a large part of the project included experimenting and designing a view that would incorporate a textual or topic analysis of texts written in a given year by a given author within the interface of the map itself. While work on this view is still in progress at the time of writing, the specification of possible view is the third portion of the essay’s results. Adding semantic context to the map visualization is particularly important to produce forms of knowledge that were not previously available to the humanist, moving beyond "at-a-glance synthesis" (to use Weingart’s phrase). As pictured in the sketch above, a possible implementation of a map tracing changes in dominant topics throughout a poet’s oeuvre over time might reimagine a node as a pie chart tracking a series of (pre-identified) salient topics generated by a topic modeling tool such as Python’s inpho on a large corpus of Russian poetic texts. In order to code values for each pie chart, partial sums would be generated from the results of the topic model, reflecting the percent of a certain writer’s work for a given year that is dominated by a topic such as "physicality", "homeland", or "Marxism".

Envisioning these topics as a pie chart would imbue the model with semantic meaning, allowing the humanist an opportunity to use the map to observe whether or not a change
in geographic location (or a permanent move away from Russia/the Soviet Union) for a given poet affected the topics they wrote about in any significant way.

But adding this extra variable of semantic context to the map visualization does not necessarily require a topic modeling of a corpus of Russian poetry. Another possible specification for charting changes in meaning on a map might simply include running a textual analysis of works written by the author in a given year using a tool like Voyant. The bubbles in this case might track the changing frequency (as a percent) of words throughout time and space, or alternatively might feature a word-map of commonly occurring words upon hovering.

There is no reason why a specification of adding a semantic component to the map visualization must be limited to textual analysis. For instance, as long as there are sufficient photos for most of the poets’ travels in a given year, it is also possible to design a map that incorporates graphic analysis (using accessible Matlab libraries) as a point of comparison over time.
Chapter 5

Conclusion and Next Steps

5.1 Conclusion

This project aimed to explore and experiment with technical possibilities for partnerships between technologists and humanities scholars, specifically applied to visual representations of Russian-Soviet émigré and traveler poets in the late Russian and early Soviet periods (1914-1941). Two of the experiments, a multimedia and a storytelling view, have been implemented and made publicly available to be used as a potential pedagogical tool for students and humanities scholars. While a third view comparing semantic features of written poetry as a function of time and geographic location was not fully implemented at the time of writing, a sketch and a few specifications for such a project were produced. Work on this essay has stressed the importance of combined efforts between technically-informed humanists and technologists with humanitarian background. Furthermore, the project shows the importance of visualizing historical and literary movements on a map not only to facilitate the display of previously available knowledge about the subjects of the map, but to carve out a path for new forms of knowledge production and cultural studies.

5.2 Next Steps: the Afterlife of Visualization

This project has been published on GitHub in an open-source repository that will continue to provide technologists and humanities scholars involved in the study of intellectual history and literary movements of the late Russian and early Soviet period. The next immediate steps are expanding upon a view that incorporates more semantic context about the poets by either mapping dominant topics throughout the poetry corpus or a crude textual analysis of individual poems.
Though the current project as implemented serves merely as a proof of concept of the importance of such visualizations, the hope is for more developers and humanities scholars to add to the initial curated data. In future versions, we hope to include a far greater number of poets that traveled farther distances than what is currently depicted. Once the number of subjects increases, building a view that depicts with the interactions (correspondences, meetings) between individual poets will be especially important. These encounters will allow scholars to see not only how geography implemented an individual poets, but will allow them to see how poets influenced one another as they interacted in that same spaces.
Bibliography


