Cartography and Code: Incorporating Automation in the Exploration of Medieval Mappamundi

Heather Wacha (wacha2@wisc.edu) and Jacob Levernier (jlevern@upenn.edu)
BH and DH Conference
University of Wisconsin, Madison
September 23, 2017

15 minutes

Introduction: (5 minutes)

HEATHER READS:

(SLIDE 2) Over the last three decades, scholars of medieval cartography have sought to reclaim the reputation of medieval mapmakers. Both the maps and their makers had often been dismissed for their perceived lack of geographical knowledge and accuracy. Recent scholarship, however, has shifted attention towards trying to understand and unravel the rationales and functions behind mapmaker’s cartographic representations in order to explore medieval conceptions of space, place and time.

(SLIDE 3) To populate their maps, medieval cartographers often combined geographical knowledge received from classical authors such as Pliny, Solinus, and Orosius, with direct experiential knowledge, (SLIDE 4) or indirect knowledge that came by way of medieval travellers and their writings. The combination of classical and medieval source material can sometimes obscure the innovative interventions that medieval mapmakers brought to their work, making it difficult to organize mappamundi into stemma, or map families. Given the approximately 900 medieval maps that
survive, we decided to explore the possibility of applying an automated, digital solution to this particularly humanistic question.

**SLIDE 5** But just to orient you first: these are two medieval mappamundi, the one on the left appears in an early 12th-century manuscript containing the annals of Peterborough Abbey, the one on the right is found in a 13th-century Psalter. If you are not familiar with medieval maps, the orientation is shifted 90 degrees to the left. So at the top of the map is the cardinal direction of east, i.e. paradise, India and parts of Asia for a medieval mind. These two maps in particular are good examples of the T-O structure - the T inside the O divides the world into the three known continents. At the top is Asia, the bottom left quadrant contains areas of Europe, and the bottom right quadrant Africa.

In order to explore the relationships between maps, traditional cartographic study relies upon methodologies drawn from history and art history, as well as geography and cartography, among others. **SLIDE 6** For example, one can make comparisons between maps based on their artistic expression, **SLIDE 7** or the organization of specific toponyms. On the Hereford map, at left, Scythia is located at the very eastern edge of Northern Asia. On the Cotton map, at right, Scythia is on the very western edge. One can also compare sets of geographical content **SLIDE 8**: on the one hand, this could mean the whole-scale organization of geographical features, such as rivers, oceans, and islands; on the other hand, this might mean comparing the corpora of
toponyms, those place names that were included in the map - as well as those that were omitted.

Until now, however, any comparison of place names between maps has been conducted by hand, usually on a case by case study. But Jacob, a data scientist, and I, a medievalist by trade, have designed a digital tool that expedites the comparison of toponyms between maps. We’ve named it veccompare (Vector Compare). It is an open-source package for the scripting language R that automates the running of “set operations” on a current dataset of over 1300 normalized medieval cartographic place names. By comparing toponyms, areas of overlap emerge between maps, as well as areas of uniqueness. It is our hope that this information will shed new light on the relationships between surviving mappamundi. At the moment, we have only eight of the 900 surviving maps in our dataset, but we’re working on it.

JACOB READS

Description of veccompare (4 minutes)

(SLIDE 9) The dataset that we have begun building comprises two elements for each map: first, we have included a diplomatic reading, which reflects the spelling of a toponym as it appears on the map. Second, because medieval Latin has no formal spelling conventions, we have taken each diplomatic transcription and “normalized” it into a version that is shared across maps. The normalized, or standardized, aspect of the dataset allows for comparisons of toponyms across maps.
This type of normalization is an editorial process -- it requires domain expertise to know whether two seemingly-related place names actually do refer to the same place. (SLIDE 10) For example, Scythia ("sith-ia") had several variants across maps. In this case, (SLIDE 11) we normalized these four variants of “Scithia” into this one spelling ("scythia"). But it was also the case that “Scythia” often had a qualifier, such as lower or upper Scythia (SLIDE 12). In this example and others like it, we decided to keep Scythia as a standalone place name, including scythia inferiore ("sith-ia inferior-āā") and scythia superiore ("sith-ia superior-āā") as separate toponyms. Noting that two maps share “scythia” and that two others share “scythia inferiore” could be more interesting to our research question, we reasoned, than noting that four maps share “scythia.” With this in mind, the dataset is constructed to clearly show the “diplomatic,” or literal, readings alongside the normalized readings; other researchers are welcome to contribute other normalization approaches for other research questions, in addition to commenting on ours.

(SLIDE 13) Additionally, we have designed and written software for facilitating the process of examining overlap in toponyms between maps. This is a package for the popular data analysis language R; a “package” in R is similar to an “add-on” or “extension” in a web browser such as Mozilla Firefox or Google Chrome -- it is freely available, can be installed by a user, and adds or makes easier new functionality for the base program. This package, “veccompare,” automates the process of comparing any number of elements, using “set operations.” Given eight maps, veccompare can automatically assess overlap and differences between all two-way comparisons of those
maps, three-way comparisons, four-way comparisons, etc. Given 100 maps, it can compute all 1-way, 2-way, etc., up to 100-way comparisons with one command. The math it does is basic -- but the package does that basic math as many times as necessary, much more quickly than a person could.

Veccompare fundamentally relies on three “set operations.” It finds (SLIDE 14) the total set of elements across maps [the “union”], (SLIDE 15) the total percent overlap across all of the maps involved in a comparison [the “intersection”] (SLIDE 16) (Here is what this looks like in a four-way map comparison), and (SLIDE 17) the percentage of map A that overlaps with map B, as well as the percentage of map B that overlaps with map A [the “relative complement” or “difference”]. (SLIDE 18) Here, for example, we see the overlap between the Munich Isidore map and the Psalter Pictorial map.

The package can produce Venn diagrams illustrating overlaps in terms of number of toponyms for up to five-way comparisons (SLIDE 19), but can perform these set operations for any number of maps. For two-way comparisons, it also automates the process of creating (SLIDE 20) overlap tables. Additionally, for two-way comparisons, it can create (SLIDE 21) network diagrams of that same overlap information. As we add more maps to the dataset, we expect that this tool will be useful for analyzing map families. It has also been designed with ease-of-use in mind: Once the package is installed, anyone using RStudio (SLIDE 22), a free and popular interface for R, can load a template that comes with the package, (SLIDE 23) point it at their data, click a button,
and (SLIDE 24) generate a report on the overlap between maps (as a website, PDF document, or Word document).

(SLIDE 25) This package is intended to be useful for and complementary to traditional manual analysis. We built veccompare for maps, but it can really assess overlap between any lists of things. In line with this flexibility, our data, and this code, are structured to support ongoing interrogation, flexible enough to adapt to shifting research questions as we progress.

In conjunction with this presentation, we are releasing both veccompare and the dataset of eight mappamundi (including the diplomatic transcriptions of the toponyms they contain alongside their normalized variants) on GitHub -- we will show the links for both again at the end of this presentation. We are hoping that others interested in this analytic approach will help and contribute to this freely accessible resource going forward.

HEATHER READS

Results (4 minutes)

This is our tool from a theoretical perspective, but how does it work in practice, and how does it help us understand book history? (SLIDE 26) For the purposes of this presentation we have chosen to focus on a historically enigmatic case: the relationship between two maps located back-to-back on the same folio in a 13-century manuscript of a Psalter. When William Bevan and Henry Phillott commented on these maps, they
were under the impression that the Psalter List map, on the right, was used as the model to construct the Psalter Pictorial map, on the left, although they did note that “the [Psalter List map] does not altogether accord with the [Psalter Pictorial map].” Even though their suggestion of model-derivative relationship makes sense, especially given that the maps share the same folio, we believe that it is unfounded. Instead of the high percentage of overlap that one would expect if the map on the left had used the map on the right as its toponymic model, (SLIDE 27) we see here that the two maps share only 44 toponyms - this is a 33% overlap for the Psalter List map and only a 26% overlap for the Psalter Pictorial map. Thus the paradox: the maps display a weak relationship vis à vis toponym overlap, yet they are inextricably linked through sheer materiality. If that physical link does underscore a strong toponymic link of some sort, it is not a direct one, and it must lie elsewhere.

Within our dataset of eight maps, the highest correspondence of toponyms for the Psalter List map is with Hugh of Saint-Victor’s *Descriptio mappamundi*, which survives not as a pictorial map, but as a narrative text that has been edited by Patrick Gautier Dalché, professor at the École Pratique des Hautes Études in Paris. (SLIDE 28) As you can see here, the overlap of place names between the Psalter List map and the *Descriptio* is much higher than those calculated between the two Psalter maps themselves. Indeed, 76% of the place names found on the Psalter List map appear in the *Descriptio Mappamundi*. (SLIDE 29) … and this percentage is somewhat lower, but still substantial, for the Psalter Pictorial map.
In order to gain perspective on how we might interpret these percentages, we can look at a comparable case that considers Hugh of Saint-Victor's *Descriptio mappamundi* and the Munich Isidore map. (SLIDE 30) Gautier Dalché argues that the Munich Isidore map is the closest surviving relative to the *Descriptio*, and goes so far as to suggest that the *Descriptio* was used as a model for the Munich map. The Munich Isidore’s 81% toponym overlap with the *Descriptio* would seem to support Gautier Dalché’s claims. Thus, in light of this one case, we might tentatively interpret that when the overlap of toponyms between two maps begins to approach 80% or higher, a “model-derivative” type of relationship between the maps is plausible and certainly merits further research. (SLIDE 31) The Psalter List map’s 76% overlap of toponyms with the *Descriptio mappamundi*, forces us to reconsider that perhaps it too, like the Munich Isidore map, used a copy of the *Descriptio mappamundi* as a model or exemplar.

The singular relationship between the Psalter Pictorial map and the *Descriptio mappamundi* appears somewhat less convincing since only 67% of its toponyms overlap. (SLIDE 32) But if our attention is momentarily shifted away from toponym overlap to the unique place names on the Psalter Pictorial map and the List map respectively - information generated by veccompare - new research questions come to light. Initially, the list of unique place names for both Psalter maps included two toponyms, Thazaron and Jazaron. Seeing the toponyms like this, side by side, highlighted their similar spelling, and once a shared geolocation was confirmed, they could then be normalized into a single place name. (Slide 33) The referent location, however, remains unclear. (Slide 34) Neither “Jazaron” nor “Thazaron” appear on any
other map in our dataset, or on any other map that Konrad Miller has annotated in his comprehensive 1897 publication, “Die Altesten Welkarten.” Miller suggests that Jazaron may reflect a corrupt version of a place name that would have been known at the time. Indeed scribal error was common in medieval manuscripts, and as more iterations were copied, more corruptions appeared. But the uniqueness of this variant corruption, found only on the two Psalter maps, suggests that the Psalter mapmakers were consulting the same corrupt source for their toponyms; and if so, Hugh of Saint-Victor’s *Descriptio mappamundi* would be prime candidate. For book historians, this kind of evidence sheds light on larger issues of manuscript production and transmission: if the maps can be linked to a common source, they can be linked to a common institution or workshop that had access to that source.

**Conclusions (3 minutes)**

**JACOB READS**

Looking toward the future, as we populate the dataset with more sets of medieval *mappamundi* toponyms, veccompare will be able to accommodate more research questions and generate more detailed results. It is our hope that map families will begin to emerge -- *(SLIDE 35)* With this in mind, one of the more surprising results from our emerging dataset was the fact that only three toponyms appear on all eight maps that we’ve looked at so far: Sicily, Ireland and Carthage.

Working on digital projects and tools such as ours has its own rewards. It has forced us to look at humanities data in new ways. As medievalists know, the data are messy.
They come from hand-written documents where spelling conventions do not necessarily exist. When “normalizing” these data, one of the contributions we can make is to document our choices and provide a rationale for them. (Slide 36) We see our contribution to the literature with this project to be the introduction of an open and documented model for collaborative data-building moving forward, as well as a new tool to facilitate and supplement research methods that humanists already practice. We think that this process -- of supplementing and automating the straightforwardly-reproducible parts of the workflow of a researcher in this field, not only provides fodder for novel, cross-disciplinary analysis approaches, but also normalizes those conversations. We thank you for listening, and look forward to your comments to help us continue to hone these data and code to become as useful as possible. (Slide 37)