PROCEEDINGS OF SPIE

Visualizing uncertainty for geographical information in the global terrorism database

Josh Jones, Remco Chang, Thomas Butkiewicz, William Ribarsky

SPIE.

Visualizing uncertainty for geographical information in the global terrorism database

Josh Jones[a], Remco Chang[b], Thomas Butkiewicz[b], William Ribarsky[b]
[a]Winthrop University, 701 Oakland Ave, Rock Hill, SC 29733;
[b]Charlotte Visualization Center, University of North Carolina at Charlotte, 9201 University City Blvd, Charlotte, NC 28223

ABSTRACT

Presenting information on a geopolitical map can offer powerful insight into a problem by leveraging an individual's innate capacity to discover patterns and to use map-related cues to incorporate pre-existing knowledge. This mode of presentation is not without its flaws, however, as the act of placing information at specific coordinates can imply a false sense of the data's geo-spatial certainty. Traditional uncertainty visualization techniques, such as those that change primitive attributes or employ animation, can create large amounts of clutter or actively distract when visualizing geo-spatially uncertain events within large datasets. To effectively identify geo-spatial trends within the Global Terrorism Database of the START Center, we have developed a novel usage of squarified treemaps that maintains the strengths of traditional map-viewing but incorporates some measure of data verity.

Keywords: Visualization, geo-spatial uncertainty; uncertainty glyphs.

1. INTRODUCTION

Data visualizations rely on certain assumptions, whether expressed or implied, regarding the quality of the data provided. Often techniques must first assume that the data is correct and then, based on this presumption, proceed to present the data in the most meaningful manner possible. There has been little progress in the design and development of data visualization techniques that can adequately present associated uncertainty values in conjunction with the data itself.

Cartography is perhaps the oldest and most frequently encountered form of data visualization that suffers from the above mentioned defect. Nevertheless, maps have continued to serve as an effective means of data analysis, synthesis, and communication. A study conducted by Stanislas Dehaene[1] indicates that the Mundurukú, an isolated indigenous tribe of the Amazon, are able to interpret maps to locate hidden objects despite having only a limited set of linguistic terms to express geometric entities and no experience with graphic symbols or geometric tools. This suggests that humans possess certain innate geometric intuitions that exist in the absence of education. In any case, this finding serves as evidence to the cross-cultural effectiveness of maps and provides further motivation for their use to communicate data and promote discovery and insight.

Presenting an object directly on a map when its geo-spatial location is only approximately known can be misleading due to the literal nature of maps. To address this deficiency, our technique augments this traditional data visualization technique by removing geo-spatially dubious entities from direct map rendering and places them, instead, in view-aware glyphs containing regionally aggregated categorical information. In doing so, this approach:

- Clearly distinguishes geo-spatially uncertain data from data that is certain

- Provides information regarding uncertain data only when the containing region is under inspection

- Provides a level of detail that is responsive to the activity of the viewer

Our paper is divided into 6 sections. In section 2 we examine the Global Terrorism Database and the particular challenges it presents. We then discuss other efforts to address geo-spatial uncertainty in section 3. In section 4 we cover the framework of coordinated visualizations of which our new technique is a part. Our new usage of squarified treemaps

is discussed in section 5 followed by a discussion of the implications of this new technique in section 6.

2. THE GLOBAL TERRORISM DATABASE (GTD)

Compiled by the Inter-University Consortium for Political and Social Research[2], the Global Terrorism Database (GTD) contains over 60,000 records of terrorist activity with as many as 119 dimensions. Due in part to the variety of sources from which this database is compiled[3], a substantial portion of these records has geographic information that is either incomplete or incorrect. This situation is further aggravated by the substantial time span covered by the GTD. Consider the twenty-one years of data concerning terrorist activity within the Soviet Union. Upon the dissolution in 1991 into its 15 constituent republics, the true location of these events becomes a challenge to determine without maintaining a separate historical database of city locations. Similar difficulties arise in cases where nations have acquired new territory or where borders have shifted over time. Further, since all locations are specified in plain English rather than cartographic coordinates, the conversion from textual descriptions to geographic coordinates introduces yet another degree of uncertainty to the dataset. Because approximately 17,000 records are affected by of the above described difficulties, the use of expert opinion to resolve these discrepancies becomes an arduous task. Our new method alleviates the need to manually clarify questionable locations by providing a geographic visualization that is capable of handling these persisting uncertainties.

3. RELATED WORK

Many techniques commonly employed to depict spatial uncertainty are ill-suited for use within map-based visualizations, especially once the number of uncertain actors becomes sufficiently large. The primary failing of many techniques suitable for spatial uncertainty depiction within Euclidian space is that they rely upon a topology that is both continuous and connected, traits lacking in the social network depicted by geopolitical maps.

The use of motion blur[4], splatting[5], and other related techniques employed to create possibility regions for the spatial location of uncertain entities are the first set of techniques that enjoy a high degree of success in Euclidian space but suffer failings within a 2D map. The first reason that these techniques cannot effectively be used within our context is because the spatial distribution produced is either uniform or distance based, neither of which can be said of human, and thus terrorism, distribution. This shortcoming is further exacerbated by the fact that many national entities are not necessarily composed of geographically connected or even proximate parts. Consider the challenge in creating a possibility region for an event that could have occurred, with equal likelihood, in either Alaska or the continental United States. Such a situation would require either two separate possibility regions, which could therefore not be addressed via a uniform or distance based distribution, or a single encompassing region that needlessly includes Canada as an area of possibility.

Primitive modification or overloading[4, 6], notable because it does not necessarily increase visual clutter, is nevertheless infeasible because it requires an initial location for geographically dubious activities. Once again, disconnected entities pose a challenge since the modified glyph can only occupy a single location at any given time. Any indication of other locations would be lost using this method.

Animation[4], also suffers from similar failings when disconnected entities are considered. These entities would require glyphs to either travel through unrelated regions or to pop between the possible locations. Such popping, although not misleading, would pose a challenge for the viewer to track, especially when multiple events within a region move concurrently. With either approach, the viewer will naturally be distracted by events of only low quality rather than be drawn to the static depictions of more reliable information.

By indicating uncertain actors through a manner that is both logically and visually separable from direct map placement, our technique is able to address disconnected, non-continuous topologies that are logically linked to physical spaces.

4. COORDINATED FRAMEWORK OF VISUALIZATIONS

In order to effectively explore the GTD, we created a coordinated framework of visualizations to simultaneously screen and report various dimensions of the GTD dataset (see Figure 1, next page). The primary visualization is a map populated with terrorism events that pass the selection criteria specified by the secondary visualizations. Once a

particular qualitative aspect of the dataset has been selected, the produced result set can be pared down by deselecting items from a legend. Further, particular time frames can be restricted and explored using a ThemeRiver (lower portion of Figure 1), a visualization technique developed by Havre et al[7] which can be used to identify categorical trends as they expand and collapse over time.

It is as part of this larger framework that our new technique realizes its full utility. Through a process outlined in the next section, we create glyphs that depict geo-spatial uncertainty which are placed upon the borders of the map view. Like each other part of the framework, when a user changes the currently selected subset of the GTD using any of the other visualizations, these glyphs react accordingly. For example, when the currently visualized set of terrorist activities is changed by the user panning the map to a new location, the border glyphs reorient themselves, optionally spawning new relevant glyphs or destroying unneeded ones, to suit the new point of view (see Figure 6). Further, when zooming in or out, as the user crosses a predefined level of detail boundary, these glyphs divide or merge to satisfy the new level of granularity (see Figure 5). Similar changes occur when the user selects new timeframes in the ThemeRiver or different categorical subsets via the legend. By using our new geo-spatial uncertainty visualization in conjunction with the framework presented here, users are able to leverage their innate capacity to discover geo-spatial patterns and incorporate pre-existing knowledge towards their analysis while remaining aware of the data's quality.
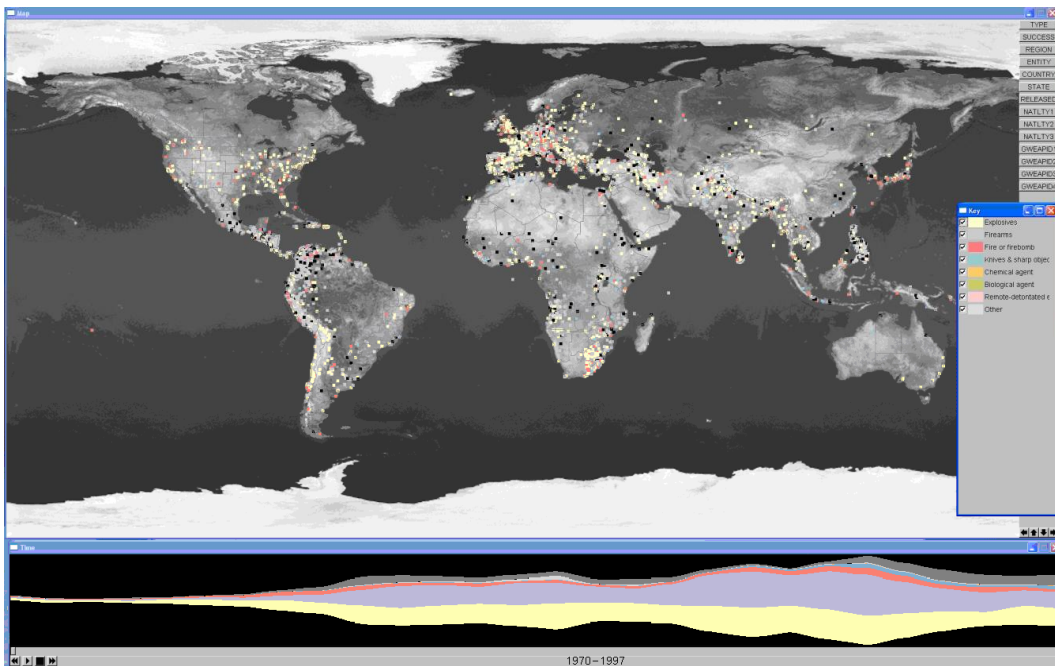

Figure 1. The coordinated framework of visualizations used to explore the GTD. The central map is the primary visualization in which terrorism events are plotted. The lower portion is the ThemeRiver, used to screen and explore temporally proximate events. On the right is the interactive key/legend can be used to exclude specific members of a categorical set.

5. SQUARIFIED TREEMAPS

5.1 Motivation

We employ a technique developed by Burls et al[8] to create glyphs known as squarified treemaps. These constructs condense the distribution of a set of records within a categorical set into a concise presentation by creating nearly square rectangles that maintain, via their relative proportional sizes, the original data distribution (see Figure 2).
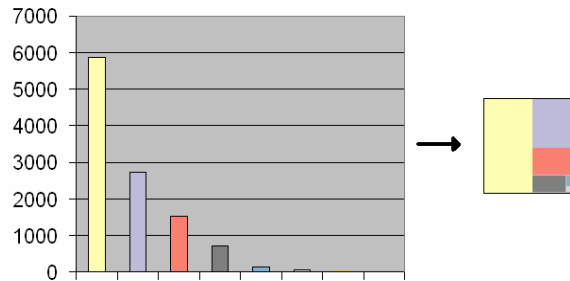
Figure 2. A squarified treemap compactly presents a categorical distribution for a set of records using a minimum of screen space, while allowing the viewer to easily observe the relative proportions.

Perhaps one of the most salient attributes of treemaps in general is their effective use of screen space to convey information, even when this space is minimal. Further, by indicating magnitude via area rather than by length, humans are better able to distinguish the relative proportions of even the smallest entries of the distribution. Squarified treemaps are superior to pie charts and standard treemaps as the latter typically produce indiscernible slivers when one or more distributions cover only a minor portion of the set under consideration. Squarified treemaps produce sub-rectangles with ratios closer to 1:1, thus avoiding thin, long slivers. This allows smaller areas to be more easily distinguishable. The nearly square entities produced are also easier to select with a mouse, facilitating user interactivity.

5.2 Squarified treemap construction

Upon program execution, a geographical hierarchy, similar to the one partially listed in Figure 3, is constructed based on the user's needs/specification. This tree can take on almost any form; the only requirement being that there exists a straightforward connection between the leaf nodes and the dataset to be visualized and that each leaf node has a defined geospatial location. We defined a hierarchy corresponding to the following geographic groupings: global, continental, regional, and national. A higher resolution level of detail could be supported given the proper data.
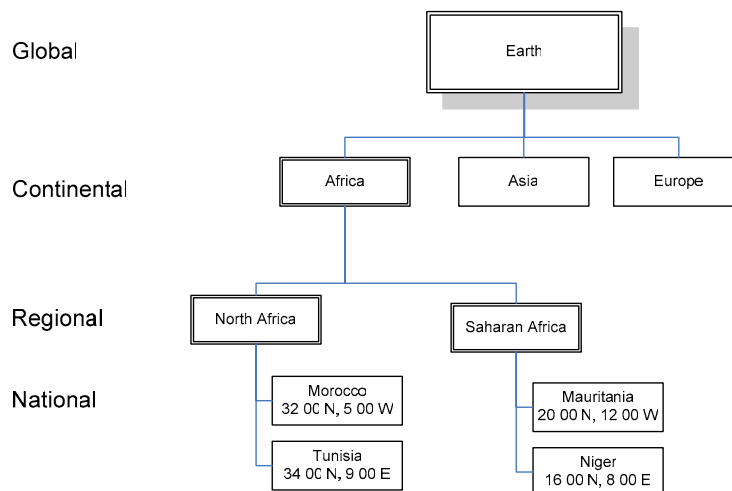

Figure 3. A partial listing of the geographical hierarchy employed to create hierarchal view-aware squarified treemaps.

Next, the coordinates of the interior nodes are constructed by taking a recursive weighted average of the descendant nodes' coordinates. Depending on the tree constructed by the user, this weighting could be based on each leaf node's land area, a user designated significance, or a simple average.

The GTD is then scanned to determine where in the hierarchy each record belongs. Leaf nodes within the hierarchy maintain a list of the record indices that could be matched but could not be assigned a specific location due to a recognized geospatial uncertainty. Interior nodes are similarly able to produce these lists by recursively concatenating the lists maintained by their descendants.

Next, the tree is traversed and a treemap container is created for each node. Although no glyphs can be produced at this time (because no categorical aspect of the GTD has been selected and thus no distribution exists), a common set of memory references are established. The first memory reference points to a "currently selected" listing of record indices that each member of the framework can modify by specifically excluding indices. The second reference refers to a list of colors to ensure consistent data labeling across all visualizations. When the category under inspection and/or the selection criteria changes, the data pointed to by the common set of memory references mentioned above is modified. The tree is then traversed, optionally updating the contents of the treemaps associated with each node of the hierarchy. Because this visualization operates in real time and the construction of treemaps is computationally expensive, nodes that are not currently visible are not updated and are simply marked as being stale. Stale treemaps defer any reconfiguration until they become visible, thus avoiding needless computation if several changes to the dataset selection criteria occur before the node becomes needed.

If a node is visible, however, the list of record indices related to the node in question is gathered and then logically merged with the "currently selected" listing. Records that pass this test are then accumulated in bins corresponding to the associated categorical value of each record. These accumulated values are then used, in conjunction with the shared coloring reference, to construct a squarified treemap that depicts the current distribution of the data under inspection.

5.3 Treemap Placement

Whenever the map display is altered, by panning or zooming actions, each node in the hierarchy is drawn to the screen using the coordinates defined for that node and is queried (using GLPicking) to determine if its associated entity is currently visible. This set of currently visible nodes is then reduced to only those existing on the current level of the geographic hierarchy under inspection. Previously visible nodes that fail have their treemaps hidden and thus only react to changing categorical selection criteria upon becoming visible. Those that are visible have their treemaps shown, reconstructing them if they are stale. The location of these nodes is calculated by projecting their location along the line extending from the center of the map, through the node's location, to the border (see figure 4). Additionally, a timer is begun to smoothly interpolate the node's movement from its previous position to its new position. This allows the user to mentally maintain the identity of a depicted treemap as it moves around the screen as illustrated in figures 5 and 6. When a zooming action causes the map to cross predefined granularity boundaries, the hierarchal depth required for a node's inclusion is modified and the above described process is repeated, resulting in a level of detail via the border glyphs. Although simple, this straightforward placement/creation/destruction process allows the border glyphs to smoothly follow the user's activity within the map, coming into and out of existence as each glyph's relevance to the task at hand fluctuates.
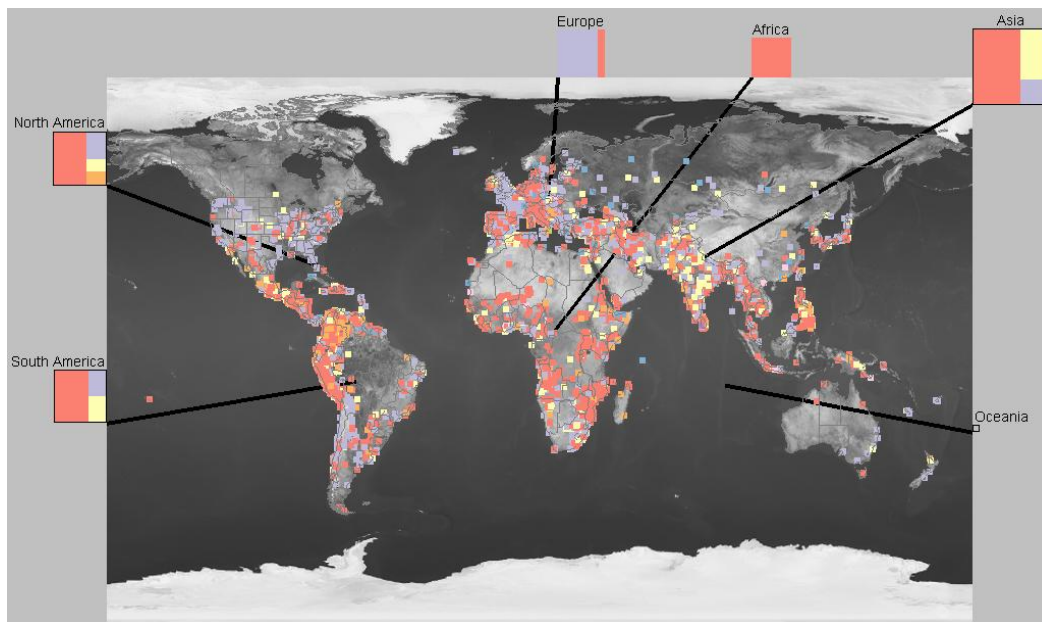

Figure 4. Uncertainty visualization has now been turned on, revealing a high level overview of the geographic distribution of uncertain records within the GTD dataset.
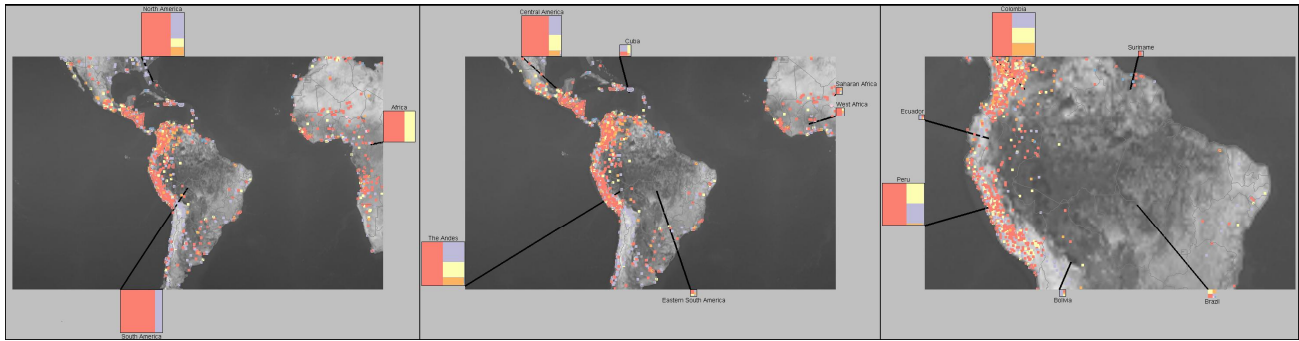
Figure 5. In response to zooming, the continental (left), regional (center), and national (right) levels of the geographic hierarchy are depicted.
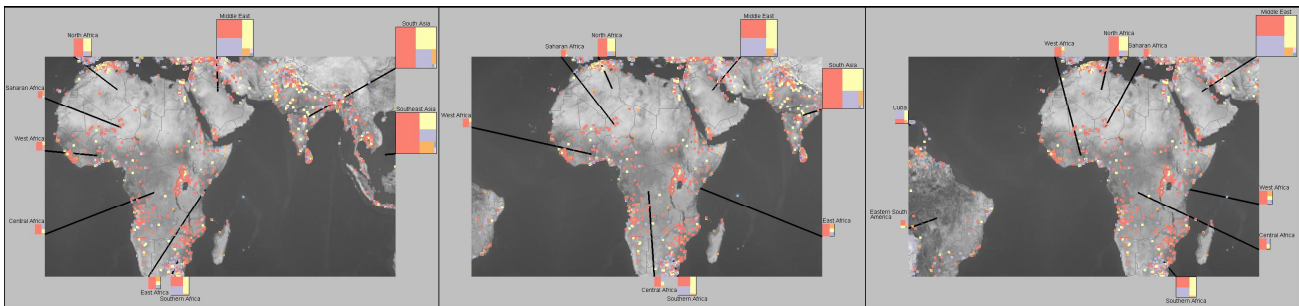

Figure 6. Panning the map causes border glyphs to reorient themselves to better fit the current perspective

5.4 Scaling

Although the process outlined above is sufficient to produce view-aware, hierarchical treemaps, one of its primary failings is that it is impossible to compare distributions across treemaps due to the absence of scale designators. Tooltips can be used to designate the particular size of any distribution, thanks to the easily selectable rectangles produced by the squarified treemap creation process. However, this is not a sufficient means for straightforward global comparison.

To address this scaling issue, whenever a treemap is shown, hidden, or the selection criteria changes, the visible treemaps are scanned to determine which contains the largest distribution. This distribution is then considered to be the standard against which all other treemaps are compared. The treemaps are then scaled to ensure that their relative screen spaces reflect their relative magnitudes.

5.5 Expanded Scope

In using the border oriented treemaps as a supplement to our primary map view, we obtained some auxiliary benefits not directly related to their primary purpose. For a dataset as expansive as the GTD, the use of any glyph, however simple, to depict an event can become cluttered once the visualized result set exceeds several hundred events. Because the squarified treemaps passively indicate aggregated information regarding the data currently under inspection, we found it useful to create a treemap associated with the *certain* data as well as the uncertain data, so that the geographic distributions of terrorist activity could be accurately compared with one another.

In doing so, we created a pair of glyphs that, when scaled, indicate not only the ratio between certain and uncertain information for a given region, but also the relative makeup for such distributions (see Figure 7). In some cases, a categorical value was completely absent from direct map placement yet existed within the geo-spatially uncertain events associated with the region, perhaps indicating a correspondence between the categorical value in question and our ability to fully document its occurrence (see Figure 8). At the very least, the use of a map without the indication provided by these uncertainty glyphs would be misleading.
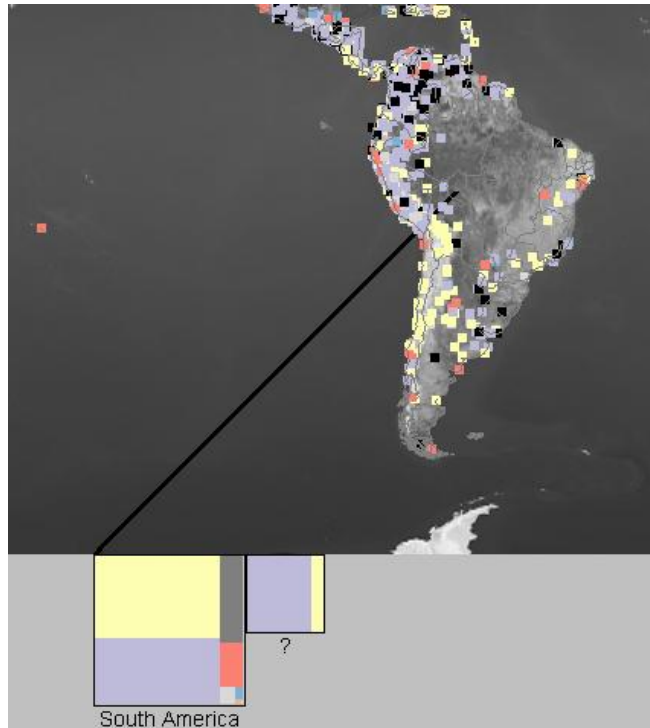
Figure 7. A quick inspection reveals that the composition of uncertain events (denoted by ?") within South America is far more homogenous than the geospatially certain ones within the same region or, indeed, of any collection at a similar level within the hierarchy. Further inspection reveals that the near equal ratio between the two leading types of attack within the certainly placed set (firearms and explosives respectively) is hardly equal in the uncertain set, and in fact is dominated by firearms rather than by explosives.
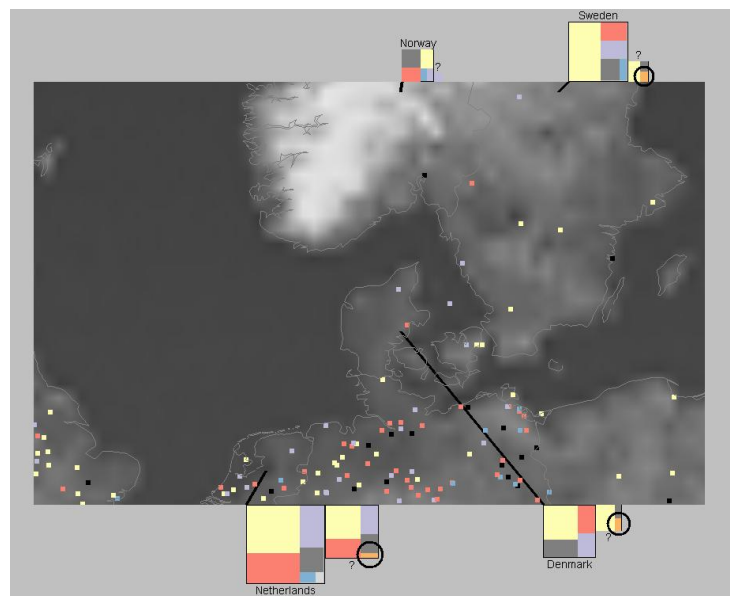

Figure 8. In Sweden, Denmark, and the Netherlands, attacks involving chemical agents (orange, circled) are completely absent from the map rendering due to inability to place theses events precisely. The use of uncertainty treemaps, however, prevents a viewer from making the false conclusion that these types of attacks do not occur in this region.

6. CONCLUSION AND DISCUSSION

Our new usage of squarified treemaps attempts to provide a means of geospatial certainty visualization that is unconstrained by the continuity or connectedness requirements imposed by alternative techniques. Because this method is connected to a definable hierarchy consisting of logically, rather than physically or proximally, grouped entities, our new technique is able to explore the GTD and similarly constructed datasets in new ways. One could, for example, replace the geo-political hierarchy used within this paper with a geo-religious one and could thus explore how the dominant religion of an area influences the activity of terrorism. The levels of such a hierarchy could then correspond to religion, branch, and denomination rather than continent, region, and nation as was used within our implimentation.

By relating these glyphs to specific locations which simultaneously correspond to differing depths within a logical hierarchy, this new approach removes extraneous data presentation while creating a level of detail responsive to the activities of the user. By presenting this information passively rather than when it is actively elicited, this technique has the capacity to draw out unsought hypothesis for an analyst to further inspect.

Squarified treemaps form the basis for our technique because they allow for quick visual comparison between all, rather than only the dominant, members of a distribution. By normalizing each treemap against a single standard, one is able to make meaningful observations without first comparing scale designators or employing overt scrutiny. Normalizing these glyphs is straightforward because they scale well, even when reduced to minimal sizes.

Although their primary purpose within our application is for uncertainty depiction, squarified treemaps often allow a user to see the global trends obscured by the large count of directly rendered events. For example, we discovered that there are stark regional differences between the target nationalities of terrorism. In Southern Europe, for example, terrorist activity typical focuses on national targets. In North Africa, however, foreign entities are normally the subject of attack.

One subject that remains to be explored is how different glyph placement strategies influence their ability to present meaningful information to the user. One limitation imposed by the border placement strategy is that the glyphs related to the center of the map are rendered furthest away from the entity they are intended to describe. Consequently, the border glyphs present a sort of peripheral vision rather than a focal one. An in-map presentation of these glyphs would alleviate this problem but faces important occlusion obstacles. Algorithms that maximize glyph proximity while minimizing the obscuring impact of these glyphs, especially in cases where crowding is an issue, need to be developed and tested.

Visualization techniques that incorporate some measure of certainty regarding the data depicted need to become the norm rather than the exception. By augmenting our coordinated framework with the technique outlined in this paper, we allow a user to make global observations regarding logically related, yet physically disjoint, geo-spatially uncertain events within an expansive geopolitical database.

REFERENCES

[1] Dehaene S., Izard V., Pica P., and Spelke E., "Core Knowledge of Geometry in an Amazonian Indigene Group," Science, 311, 381-384 (2006).

[2] LaFree, G. and Dugan L., GLOBAL TERRORISM DATABASE, 1970-1997. ICPSR04586-v1. College Park, MD: University of Maryland, 2006. Ann Arbor, MI: Inter-university Consortium for Political and Social Research, 2007-04-04.

[3] Dugan L., Lafree G., Fogg H., [Intelligence and Security Informatics], Springer, Berlin & Heidelberg, 407-419 (2006).

[4] Pang A., Wittenbrink C.M., Lodha S.K.. "Approaches to uncertainty visualization," The Visual Computer, 13(8), 370-390 (1997).

[5] Rheingans P. and Joshi S., "Visualization of molecules with positional uncertainty," Data Visualization '99, 299-306 (1999).

[6] Rhodes P. J., Laramee R. S., Bergeron R. D., and Sparr T. M., "Uncertainty Visualization Methods in Isosurface Rendering," Eurographics 2003, 83-88 (2003).

[7] Havre S., Hetzler E., Whitney P., Nowell, L., "ThemeRiver: visualizing thematic changes in large document collections" IEEE Transactions on Visualization and Computer Graphics, 8(1),9-20 (2002).

[8] Bruls M., Huizing K., and van Wijk J. J., "Squarified Treemaps," Proc. of Joint Eurographics and IEEE TCVG Symp. on Visualization (TCVG 2000), 33-42 (2000).