Geo-Storylines: Integrating Maps into Storyline Visualizations

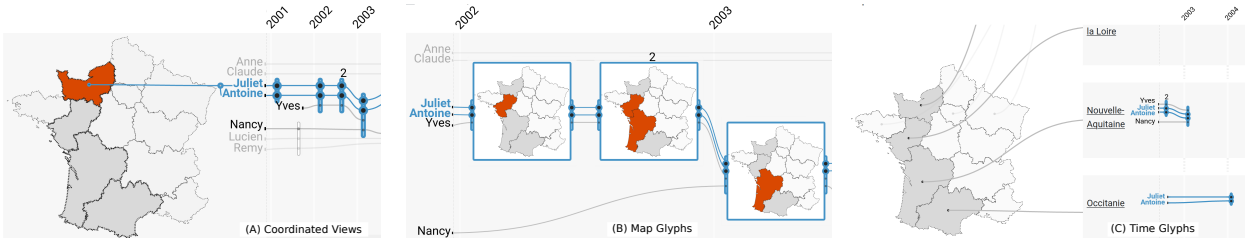Golina Hulstein, Vanessa Peña-Araya, and Anastasia Bezerianos

Fig. 1: Three Geo-Storyline designs showing the geo-temporal evolution of the relationships between people. The two selected people in blue moved together from north to south. (A) **Coordinated Views** include a map on the left and a unique Storyline timeline on the right. While scrolling, links appear between the relationship nearest to the map and the associated locations. (B) In **Map Glyphs** each relationship is represented by a map with the associated locations drawn in orange. (C) **Time Glyphs** are composed by a map on the left and a scrollable list of Storyline glyphs on the right. Each Storyline glyph contains all the relationships associated with one location. Gray lines between a location on the map and a storyline glyph indicate the correspondence between the two.

Abstract— Storyline visualizations are a powerful way to compactly visualize how the relationships between people evolve over time. Real-world relationships often also involve space, for example the cities that two political rivals visited together or alone over the years. By default, Storyline visualizations only show implicitly geospatial co-occurrence between people (drawn as lines), by bringing their lines together. Even the few designs that do explicitly show geographic locations only do so in abstract ways (*e.g.,* annotations) and do not communicate geospatial information, such as the direction or extent of their political campains. We introduce Geo-Storylines, a collection of visualisation designs that integrate geospatial context into Storyline visualizations, using different strategies for compositing time and space. Our contribution is twofold. First, we present the results of a sketching workshop with **11** participants, that we used to derive a design space for integrating maps into Storylines. Second, by analyzing the strengths and weaknesses of the potential designs of the design space in terms of legibility and ability to scale to multiple relationships, we extract the three most promising: Time Glyphs, Coordinated Views, and Map Glyphs. We compare these three techniques first in a controlled study with 18 participants, under five different geospatial tasks and two maps of different complexity. We additionally collected informal feedback about their usefulness from domain experts in data journalism. Our results indicate that, as expected, detailed performance depends on the task. Nevertheless, Coordinated Views remain a highly effective and preferred technique across the board.

Index Terms—Storyline visualization, geo-temporal data, maps, hypergraphs

---

1 INTRODUCTION

In 2009 the XKCD comic introduced narrative charts, a visualization to summarize movie plots [35]. In these charts, each character in the movie is represented by a line along an implicit horizontal timeline. Characters' lines curve towards each other when they meet, stay close as long as they are together, and move apart again when they split up.

This co-occurrence between movie characters can be more broadly defined as a relationship among entities of any type that occurs on a specific point in time. Their succinct layout makes narrative charts, later known as *Storylines*, a powerful tool applicable to a variety of domains. For example, they have been used to visualize how developers interact on a code repository [36], to illustrate spatio-temporal patterns of fixations in eye tracking [9], and to visualize the evolution of communities in dynamic social networks [44]. More recently, HyperStorylines has generalized Storylines to represent complex relationships between any type of entity in a hypergraph [41].

- *Golina Hulstein is with Université Paris-Saclay. E-mail: golina@hulstein.com*
- *Vanessa Peña-Araya is with Université Paris-Saclay, CNRS, Inria. E-mail: vanessa.pena-araya@inria.fr*
- *Anastasia Bezerianos is with Université Paris-Saclay, CNRS, Inria. E-mail: anastasia.bezerianos@universite-paris-saclay.fr*

Relationships among people in real world contexts commonly include geographical locations (*e.g.,* countries or cities) or other contextual places (*e.g.,* a cafe or a theater). Let us consider the example of our journalist experts analysing the election campaign of two political rivals in France, Juliet and Antoine, shown in Fig. 1. We can define each political debate between them as a relationship, linking the two candidates, a city and a date. A more complex relationship can consider other events organised by their parties occurring simultaneously in different cities. Here, a single relationship can link the names of the organizers, several locations and a date. Journalists may investigate the structure and impact of these campaigns, trying to understand the cities the politicians visited together, if the supporting events happened between contiguous cities or were more wide-spread, and whether the campaign events were organized simultaneously or moved sequentially from city to city following a geographical pattern. Answering these questions requires a visualization of geospatial, temporal, and relationship data. For example, in Fig. 1 we observe that the relationships that link Juliet and Antoine start in the north of the country in 2001 and move slowly to the south over the years.

Some Storyline variations use abstract strategies, such as line position or annotations, to visualize where relationships occur (*e.g.,* [7, 31, 50]). However, they lack geospatial detail and therefore cannot answer all of the questions mentioned above. In addition, they do not consider complex spatial links where more than one location is linked to a single relationship. We address this gap by introducing and studying *geo-storylines*, visualizations that integrate the full geospatial context of relationships into storyline visualizations.

Our work makes the following two contributions:

1. The Geo-Storylines design space, derived from sketching workshops and a literature review of tangentially related visualizations.

2. The evaluation of the three most promising designs from the design space. First, in a controlled user-study that focuses on five spatio-temporal tasks: identify locations, identify dates, distance, direction and adjacency. And second, through expert feedback. The three designs differ in their focus on space vs. time: **Map Glyphs** split spatial information over multiple maps within the storyline; **Coordinated Views** link the temporal information in the storyline with a separate map on user interactions; and **Time Glyphs** split time by representing it in a separate storyline for each geospatial location.

Our results found Coordinated Views more efficient and preferred overall. However, both Map Glyphs and Time Glyphs had higher accuracy than Coordinated Views when identifying locations. Our code, supp.material and link to a live demo are available at `https://gitlab.inria.fr/ilda/geo-storylines`, and experimental analysis and scripts can be found at `https://osf.io/5wnyg`.

2 RELATED WORK

In this section we first summarize literature on the visualization of relationships over space and time in generally. We then cover existing work on Storyline visualization, and relevant design spaces.

2.1 Visualizing Relationships over Space and Time

There is an extensive amount of research on spatio-temporal visualizations. For instance, there are general approaches to represent change over space and time, like small multiples [20, 30, 56], animation [32] or 3D views [27, 52]. Other approaches focus on specific types of data, like flow of movement [45, 56], trajectories [8, 53, 57] and other spatio-temporal patterns [34]. Nevertheless, this work does not include information about the evolving spatio-temporal relationships among entities, such as when the entities come together or diverge in their trajectories. When we do not consider time, there is also considerable work expressing the relationships of only two entities, expressed as nodes, in the form for geospatial networks. This work is summarized in the survey by Schöttler *et al.* [47].

There is little work that considers multiple entities (*e.g.,* people), whose relationships evolve in space and time. Examples of such data can include genealogies, academic co-authorship networks, or political rivalries, and can be modeled as *dynamic hypergraphs*. Hypergraphs are the generalization of graphs, where links can include any number of nodes instead of only two [11], and nodes can be of any type (*e.g.,* people, locations). The visualization of dynamic hypergraphs, that represent the evolution of relationships over time, is an active research field [22]. The three more recent visualizations are PAOHViz [55], Hyper-Matrix [21] and HyperStorylines [41]. The first two adopt a matrix approach and represent entities as rows and relationships in columns. HyperStorylines, on the other hand, generalizes Storyline visualizations to show relationships between multiple types of entities. HyperStorylines and PAOHViz both encode location entities, but these are treated as any other named entity without explicit geographical context to allow for geospatial analysis (*e.g.,* the distance or spread between the locations where two relationships occur).

In summary, there are no existing visualizations that describe the geospatial context of entities' relationship dynamics. Furthermore, most of geospatial network visualizations do not focus on links of more than two elements. In this paper we set out to create visualizations that can combine entities' dynamic relationships with the geospatial context they occur in. We chose to do this by integrating geospatial information into Storylines as their layout prioritizes grouping of related entities and allows for a more compact and easy-to-analyze visualization than PAO-HVis [41]. It is this relative simplicity of Storylines that makes them a promising candidate for integrating additional geospatial information.

2.2 Storyline visualizations

Inspired by the XKCD's Narrative Charts, Ogawa and Ma introduced Storyline visualizations, including a layout algorithm, that was used to display interactions between software developers [36]. Later research has mostly focused on their automatic generation [7, 24, 31, 33, 48, 49]

and on applying them to different domains. For example, Storylines have been used to visualize dynamic social networks [44, 59], collaboration in groups [33, 36], genealogical data [28], temperature changes over time [58], and even to analyze eye tracking data [9]. A more recent line of research considers how additional information can be integrated into Storyline visualizations including mixing automatic and human input [50, 51], non-linear narratives [37, 38] and multiple relationships at once by branching their lines [17].

Only a handful of papers consider space in Storylines. The original Narrative Charts [35] use labeled background contours to encode space, which was later extended with layered contours that can show location hierarchy in a form similar to a river plot [31, 48, 49]. Other visualizations use arc segments drawn between entities [43], or the y-axis position of entities to encode space [7]. Finally, some hand-drawn Storyline examples use line rendering properties to show transitions between worlds [50, 51].

The aforementioned representations of space show the geospatial context to a limited extent. Layered background contours in the form of river plots provide information about the hierarchical relationships between locations, but do not show information related to distance or direction. Designs that reduce 2D space to the y-axis can show distance but distort the horizontal space making horizontal direction or adjacency hard to determine. One exception is the work of Yagi *et al.* [58] that shows temperature variation in Japan. Their visualization combines a map of Japan with a Storyline representation of temperature, and relates the two through color. While this design does describe the full geospatial context, its reliance on color strongly limits the number of locations it can encode.

2.3 Existing Design Spaces relevant to Storylines

Some existing design spaces are relevant to Storylines, although they do not necessarily consider geographical context. For example, Tang *et al.* [51] created a design space for their narrative elements. Their design space is structured into five dimensions describing ways to encode actors, relationships, and plots as well as variations of timeline structure and decorative elements. However, their design space does not cover geospatial representations in Storylines.

Brehmer *et al.* created a design space for storytelling with timelines [13] which is structured into three dimensions: how a timeline is represented (linear, radial, a grid, etc.), the scale used in the timeline, and its layout. It does not, however, address Storylines specifically.

Bach *et al.*'s recent review of geospatial network visualizations provides a comprehensive overview of how networks can be combined with geospatial visualizations [46]. Their design space describes the representations of space and networks, but also considers how the two are combined. Although this design space does cover geospatial representations, it is missing the time dimension. On the other hand, the more general framework defined by Bach *et al.* [8] gives a descriptive model for temporal data. However, it does not consider explicit relationships among entities, nor geo-spatial data. Finally, the framework by Windhager *et al.* [57] visualizes biographical data over space and time using multiple 3D views. It describes people's trajectories, but does not explicitly represent relationships between them.

While none of these design spaces describe how geospatial context can be integrated into Storylines, they provide inspiration for the dimensions proposed in our own design space and are discussed there.

3 GENERATIVE WORKSHOPS

The challenge of analysing the spatio-temporal properties of evolving relationships is important in several domains, including data journalism. The example provided in the introduction comes from a long collaboration with data journalists [41] (see Sect. 7), where we identified the need to add topological / geographical information in Storylines. We thus set out to explore the different forms that Geo-Storyline visualizations could take. As existing approaches include limited spatial information in Storylines, we decided to generate a breadth of alternative visualizations to better understand their properties and limitations. Following an approach similar to Tang *et al.* [51], we conducted design workshops where participants generated hand-drawn visualizations, which we then

used to create a design space. We describe our workshop considerations, procedure and the analysis of the created designs.

3.1 Spatial relationship patterns and tasks

We asked participants to generate designs that combine three dimensions: people, and how their relationships evolve over space, and over time. As we are interested in adding geo-spatial context, we identified early on five patterns that describe how people's locations relate to one another: for example when two politicians visit cities which direction do they follow, or what distance do their respective campaigns cover. These patterns are: distance, direction, extent, adjacency, and hierarchy (discussed next). In the workshop, we used these patterns to inspire participants to include different aspects of space in their sketches (Sect. 3.3). We then used them as a basis for the tasks in our comparison study (Sect. 5). We next explain why we chose these patterns and how we turned them into questions and tasks.

To obtain the geospatial patterns of how locations relate to each other, we looked at the taxonomy of movement by Dodge *et al.* [18], that organizes the possible parameters that define a movement pattern. We focused on their primary derivative parameters that define basic movement. Additionally, we do not include parameters in the temporal dimension (*duration*) as we focus on the spatial context of relationships. This led us to the spatial patterns: *distance*, *direction*, and *spatial extent*. When used as prompts, participants in the workshop would be asked to design a visualization that could answer questions such as *what distance did these two people travel together?*, *which direction did they take?*, and *how much of the country did they cover?*

We consider two additional spatial relationships that stem from the structure of maps: *adjacency* and *hierarchy*. Similarly to work on visualizing spatial propagation [39], we consider two locations as adjacent when the trajectory between them is contiguous and is not separated by other locations (no hops). Geo-spatial hierarchy describes the common spatial organization of map locations into levels such as cities, countries, and continents. Prompt questions for these relationships would be *were there any jumps in the trajectory followed by these two people?* or *how did people move between countries and continents?*

3.2 Participants and data

We conducted three 2h workshops with 11 participants from our institution (7 women, 4 men, 1 did not specify). Workshops were divided in two 1h sessions and were conducted online. In an effort to generate usable and varied designs, all participants were practitioners or researchers in HCI and visualization.

Participants could choose to prepare their own dataset or use one of two prefabricated datasets: (i) the movements of a subset of Lord of the Rings (LotR) characters, and (ii) a music tour of Imagine Dragons between 09-2017 and 08-2018 with its opening acts. For both datasets we provided a map. Five participants chose to use the LotR dataset and two chose the Imagine Dragons one. The remaining four participants chose to collect a personal dataset about their trips with family or friends, and were instructed to collect data about when and where the relationships between people occurred. To keep the datasets manageable in size for creating hand-drawn sketches, we limited them to 3-5 people and approximately 15 relationships.

3.3 Procedure

The workshop was structured into two sessions of one hour each. In the first session, we introduced the theme and showed participants a set of basic visualizations of time, space, and relationships as inspiration (slides provided in supp.material). We note that storylines were only one among several inspiration visualizations. Participants were then given 25 minutes to sketch a first, unguided, visualization of their data. They were asked to upload their sketches to a Miro board[1] and then present them to the group. At the end they were asked what remaining questions they had about their data.

Before the second session, we prepared two sets of 2-3 questions for each participant to choose from. Allowing participants to choose from

---

[1] www.miro.com

a set of constraints has been used in previous visualization workshops to foster creativity in a reasonable time span [26]. The first set came from the questions participants reported having about their data at the end of session 1. The second set asked participants to focus on answering a question related to one of the identified geospatial patterns (see Sect. 3.1), guiding participants to explore different aspects of space in their sketches such as distance, direction, spatial extent, and hierarchy. The second session started with a short intro, after which participants completed two rounds of sketching of 20 minutes each. In each round they chose one of the questions we prepared. We also gave participants a set of inspiration cards (supp. material) to remind them of different ways to visualize space, time and people.

As our goal was to generate a wide variation of visualizations, all three authors participated in the sketching during the workshops. We used the Imagine Dragons dataset as well as our own custom datasets.

3.4 Analysis of Generated Sketches

The workshop produced 67 sketches (supp.material), among which 39 were created by participants and the rest by the authors. However, not all sketches visualize multiple people and their relationships over space and time. For example, 5 participant sketches visualized the movements of only one person, failing to describe relationships. And notably, 11 sketches (7 from participants) are missing the time dimension. Based on our observations, this happened when participants made the strategic decision to start their visualization with only two aspects of the data (*e.g.,* people and space). In some cases, they were able to integrate the third aspect (*e.g.,* time) later on, but sometimes they failed to do so indicating that combining all three aspects is challenging. Indeed, at the end of the workshop several participants expressed how difficult it was to combine all three dimensions.

It is interesting to note that while participants were provided with several geo-temporal inspiration visualisations (not only storylines) they often naturally converged to sketches that have storyline visualization characteristics (21 out of 39 generated by participants). Examples of sketches can be seen on the left and right side of Fig. 2.

We performed a bottom-up card sorting of the workshop sketches in several iterations, grouping them by visual encoding, and by what aspects of the data they highlight. We examined the resulting clusters in terms on whether they can represent the relationships of multiple people and their spatial evolution. We eliminated clusters that rely on abstract representations of space, as they are either unable to show spatial relationships between locations, or only show a single spatial relationship (such as distance or hierarchy only). This resulted in five clusters whose characteristics we analysed to create a design space (Sect. 4), made up of 41 sketches (25 from participants). Details of our iterative analysis, resulting clusters and all sketches are in supp.material.

We note that the workshop sketches did not always include Storylines: for example some included other timeline representations, like Fig. 2H that associates a separate custom timeline representation to each location. We thus created representative designs for each cluster that illustrate the sketch concepts, but are adapted for storylines. For example in Fig. 2H we replaced the custom timeline with a Storyline. These representative designs are seen at the center of Fig. 2. As the *Small Multiples* cluster presents two promising design variations, one with a map per relationship, and one with a map per time-step, we included two representative designs, seen in Fig. 2A,E respectively.

4 THE DESIGN SPACE

Here we introduce the Geo-Storylines design space, that considers how to best incorporate geographical context into storylines. We first describe its three dimensions that explain how space and time can be combined, how the design space categorizes the designs extracted from the workshop, and how it was used to generate two additional designs. The design space, along with representative designs and inspiration sketches, can be seen in Fig. 2.

4.1 Design Dimensions

We observed that the six representative designs extracted from the sketching workshop (Fig. 2) may split either the space or time rep-
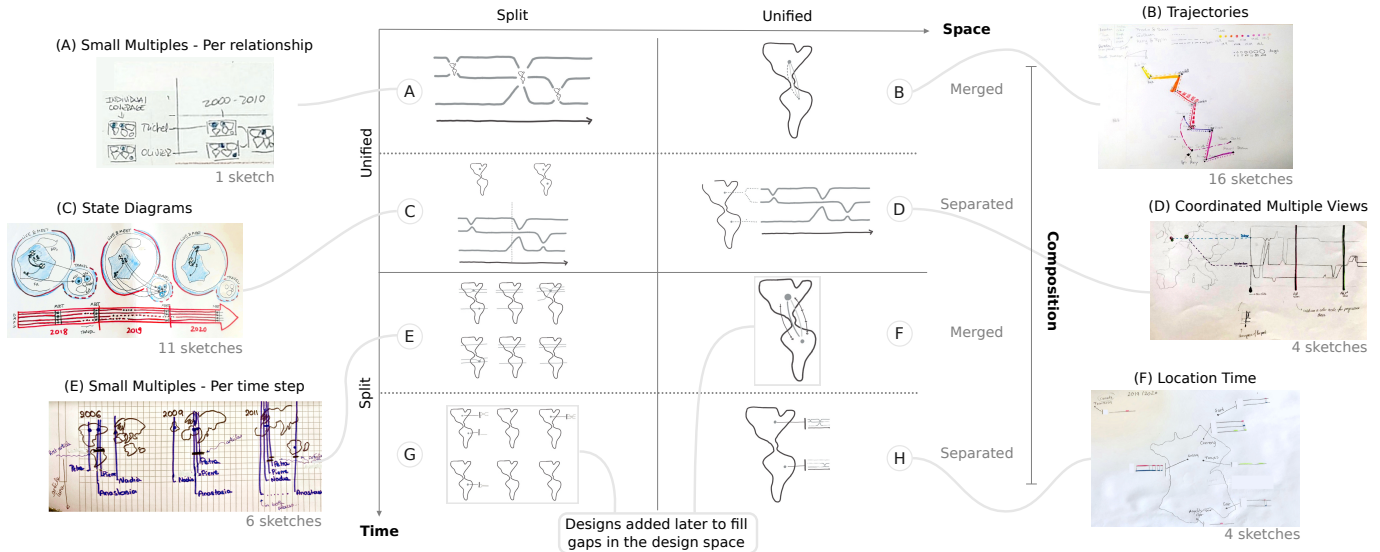
Fig. 2: Design space. The central part of the image provides an overview of the design space, with the six representative designs selected from the analysis of our sketching workshop (A-E,H). The two representative designs indicated with a gray border (F,G) were added later to fill gaps in the design space. Surrounding the design space are sketches from each of the clusters identified in the workshop, that inspired the different representative designs. The cluster names are indicated above the sketches, and the number of sketches in them below in gray.

resentations in the layout, depending on their emphasis (keeping the dimension of emphasis unified). This difference between *split* and *unified* layout representations of time and space defines the first two dimensions of our design space. However, there are cases where these two dimensions were insufficient to differentiate all the potential designs. For example, Fig. 2A and Fig. 2C both maintain time unified and space split, yet they clearly differ in how they combine these two dimensions. This difference motivated the third dimension: the overall composition of the visualization, which can be *merged* or *separated*.

We note that the design space resulting from these dimensions had two cells with no representative design, so we generated two additional designs to fill the gaps, seen in Fig. 2G and Fig. 2F. We next discuss the characteristics, advantages and disadvantages of these dimensions.

Time Layout. The layout of time can be either *unified* (a single timeline) or *split* (one timeline per location). Our time layout dimension resembles the one proposed in the design space for timelines by Brehmer *et al.* [13]. They distinguish between unified, faceted and segmented layout. The latter two, where time is split by a categorical value (faceted) or temporal divisions (segmented), correspond to our split time layout. When considering the Space-Time Cubed framework for temporal data by Bach *et al.* [8], we can classify Storylines as a space flattening operation and our design space as chained operations. However, this framework does not address space explicitly, and thus does not acknowledge the way space is embedded (*e.g.,* in Fig. 2A).

For *unified time*, Fig. 2 shows three designs that use a single linear timeline, with a map positioned either: inside it (Fig. 2A), on top (Fig. 2C), or next to it (Fig. 2D). In the case of (Fig. 2D), links between the map and the timeline appear interactively by scrolling horizontally. In a fourth design (Fig. 2B) time is encoded on trajectories: people are differentiated through line symbols (dashes, dots, etc.) and time by a common time-color gradient (higher saturation indicates recency).

For *split time*, Fig. 2 shows how we can split time: by periods (small map multiples in Fig. 2E and G), by locations with one timeline per location (Fig. 2H), or by people and individual trajectories (Fig. 2F).

Space Layout. Similar to time, the layout of space can be either *unified* (one map) or *split* (small multiples). The spatial layout dimension resembles Peña-Araya *et al.*'s categorization of spatio-temporal visualizations [40]. In their categorization, visualizations are classified based on whether they juxtapose time (*e.g.,* animation) or locations (*e.g.,* small multiples). While their categorization strongly overlaps with the space layout dimension in our design space, it does not cover

all cases in the Geo-Storylines design space. Notably, one of our Geo-Storylines designs juxtaposes space based on relationships rather than based on time (Fig. 2A).

For *split space* we have designs that use one map per relationship (Fig. 2A) or per periods of time (Fig. 2C, E and G). For *unified space*, we see one map with merged or split timelines inside it (Fig. 2B,F), or next to the map (Fig. 2D,H).

Composition. The third and final dimension describes how the representations of time and space are combined into one visualization. Time and space representations can be either placed side by side - keeping them *separated* - or they can be *merged*. The composition dimension also exists in Schöttler *et al.*'s survey of geospatial networks [46]. They define four categories of compositions ranging from loosely to strongly integrated: juxtaposed, superimposed, nested, and integrated. Superimposing, nesting and integrating all describe different ways to merge the representations of time and space, and are all three present in the Geo-Storyline designs. In one design, small maps are nested into the storylines for each relationship (Fig. 2A). In other designs, time is integrated into trajectories (Fig. 2B,F). Finally, one design superimposes timelines over maps (Fig. 2E). As we have only one or two examples using superimposed, nested, or integrated composition each, we combine these categories into *merged* compositions.

4.2 Design Assessment

We assessed the capabilities of each visualization in an effort to identify the most promising designs. Our goal was to identify designs that represent all individual relationships without overlapping and without requiring arbitrary aggregations, *i.e.* designs that show all available relationship information. In particular, avoiding aggregation ensures that different relationships occurring at the same time and place are distinguishable.

The designs Fig. 2B and F are incapable of differentiating multiple relationships occurring at the same time and location(s). These designs, as well as Fig. 2E also suffer from occlusion. All three place relationships as glyphs over a map, so the available space to position a set of relationships is constrained by the boundaries of the location they are associated with. Small locations or those with large number of relationships will inevitable suffer from occlusion.

Design Fig. 2C aggregates several relationships over time, shown together on one of the mini-maps on top of the timeline. This aggregation makes it difficult to differentiate relationships happening in the same location at different times. Extending the design to one map per

relationship would eliminate the problem. However, it would require users to split their attention between navigating the storyline and identifying the corresponding mini-map on top. This makes it a less legible version of (A) where the mini-maps follow the people's lines.

Design Fig. 2G also requires aggregation of time. And as it combines multiple time glyphs with map glyphs, it subdivides time and space to such an extent that it becomes a less illegible variation of Fig. 2H.

Three promising designs remain: the small multiple maps per relationships that we will call from now on **Map Glyphs** (Fig. 2A); the design where each location has its own storyline, called from now on **Time Glyphs** (Fig. 2H), and the **Coordinated Views** (Fig. 2D).

5 USER STUDY

In this section we describe our evaluation of the three selected Geo-Storylines designs. The study aims to evaluate two aspects of the designs: 1) How effectively they visualize the geospatial relationships described in Sect. 3.1; and 2) how their performance scales with the number of locations they visualize. Figure 1 shows an example of the implementation of the three selected designs. This dataset is inspired by a real example from data journalism. It contains ten relationships between seven people, at six different points in time, meeting at four locations on a map. In each visualization, two people are selected.

5.1 Implementation of the designs

For our study, we implemented the three Geo-Storylines designs as an online interactive system. We took care to make interactions and the visual representation of relationships and entities (time, location, people) as similar as possible across the three visualizations.

Visual Properties: Designs vary in how to represent the links between time and geography. However, all timelines follow a basic Storyline visualization where each line represents a person that evolves over time in the horizontal axis. Lines, representing people, come together inside relationship boxes each time there is an event that involves all these people. Relationship boxes are rounded rectangles, with a hinge (dot) at the point where peoples' lines go through them. In the case of Map Glyphs, the relationship box is replaced by a mini-map. All locations of the map are colored with gray at the beginning. We limited the use of color as much as possible to push the basic visual encoding of each design. Each visualization fits in 2564×1310 pixels.

Interactions: Our design space does not consider interactions (except scrolling for Coordinated Views). When rendering real world datasets (see Sect. 5.5) it became clear that realistic timelines, where the names of the entities are big enough to be readable, extend beyond the available screen real-estate. So basic scrolling was added to all visualizations.

In addition, we observed in a pilot that visually tracing the path of specific entities while scrolling was tedious, as entity lines often cross and curve. We thus added basic selection functionality, that also exists in many other Storylines systems (*e.g.,* [5, 36, 41]) The three designs allow users to select one or more people, locations or dates. Selected entities are highlighted in blue, and so are the borders of the relationship boxes they are involved in. When more than one entity is selected, it is the relationship boxes that include *all* these entities that get highlighted in blue (intersection). Once a selection is activated, the remaining entities are faded out: if they are connected with some (but not all) of the selected entities they become gray, and if they are not connected to any selection they fade out to a light gray. This distinction allows users to differentiate between related, but unselected entities (gray) that may provide contextual information, and ones they do not relate to user's selection at all (light gray). We did not add advanced interaction, such as filtering or zooming, as we wanted to study the core design trade-offs of the techniques, distilled in our hypothesis in Sect. 5.3 (*e.g.,* Time Glyphs require vertical space for all timelines, Map Glyphs result in small resolution mini-maps).

We next note differences across the designs (also seen in Fig. 1).

Coordinated Views (Fig. 1A): Coordinated Views contain a map on the left, and on the right a single Storyline visualization. To see which locations are associated with a relationship, users scroll until the relationship's box is close to the left border of the timeline (dashed vertical line). Map locations that are associated with the upcoming relationship

in the scrolled storyline, are visually linked with the relationship box and colored orange (to differentiate them from blue selections).

Map Glyphs (Fig. 1B): In this case, the base Storyline visualization stays the same except that each relationship box is replaced by an embedded mini map. All locations that are part of the relationship are colored with orange. Users can select any location inside the mini maps and the selection will be reflected in all the mini maps of the timeline.

Time Glyphs (Fig. 1C): A map is positioned at the left. On the right, there is a scrollable vertical list of multiple Storylines glyphs, one per location. Storyline glyphs are linked to their associated location in the map with a gray link. Users can scroll vertically to see the different storylines. Each time a location is selected on the map, an animation puts the corresponding storyline in the center of the view.

5.2 Tasks

Similarly to HyperStorylines [41], we considered the Andrienko & Andrienko task taxonomy for spatial and temporal data [6]. Our goal was to include tasks from both reading levels of this taxonomy: *elementary* (about individual elements of the dataset) and *synoptic* (that require looking at the entire dataset or a subset of it). We started with the spatial relationships used in the workshop (distance, direction, adjacency, coverage, hierarchy - Sect. 3.1), which can be categorized as *synoptic*. To include elementary tasks, we added two baseline tasks, inspired by the basic tasks used to evaluate PAOHViz [55] and HyperStorylines [41]: "find all relationships between person A and person B". In our case, this requires users to find either all the locations or all the dates involved in the relationships between a set of people.

Finally, after an initial pilot we realized that the strategies to perform the tasks *coverage* and *hierarchy* were very similar to the baseline task of identifying all locations involved in a set of relationships. Therefore, we only kept the simpler task of identifying locations (next described as Baseline Locations) to reduce the duration of the study. We thus tested the following tasks:

- [T1] Baseline Locations: *"In how many different locations did person A and B meet?"* (numeric answer). This task is an *elementary look-up* as it requires participants to find a set of individual targets (locations) given a set of references (people).

- [T2] Baseline Time : *"On how many different dates did person A and B met in location L?"* (numeric answer). Similar to the previous one, this is also a *elementary look-up* task.

- [T3] Distance : *"In which location did person A and B meet that is closest to location L?"* (answer is a location's name). This is a *elementary comparison* task as it requires participants to compare the distance of a set of individual locations in relation to an specified one.

- [T4] Direction: *"Which direction did person A and B follow when they met?"* The answer to this task can be: north-south, south-north, west-east, east-west, or no pattern. This is a *synoptic relation-seeking* task that requires participants to see the overall trajectory two people followed.

- [T5] Adjacency: *"Did person A and B move only between adjacent locations?"* Binary answer between "Yes, they move only between adjacent locations" and "No, there one or more jumps in their path". This is a *synoptic relation-seeking* task, as it requires participants to observe the overall pattern and check for neighboring locations linked to the two central people across time steps.

We explain why we focus on relationships between two people in the dataset creation method (Sect. 5.5).

5.3 Hypotheses

We describe next our formulated hypotheses for our user study and the reasoning behind them:

[H1] Movement patterns - *CoordV > MapG > TimeG*
Previous research shows that animations are better suited to identifying sudden changes in movements [14, 23, 39], followed by small-multiples. Therefore, for the tasks that track movement (Direction and Adjacency) Coordinated Views should perform best, as it is essentially a user-controlled animation. Map Glyphs will follow as they behave similarly to small multiples. Finally, as Time Glyphs split

time by location we expect participants will have trouble following the movement of people across several split timelines.

[H2] Identifying locations - *CoordV & TimeG > MapG*
Designs with a single unified representation of space are better suited for tasks that require identifying all the locations linked to a group of people. We thus expect Coordinated Views and Time Glyphs to perform better in `Distance` and `Baseline Locations`. For Map Glyphs, the selection interaction may mitigate some of the challenges when searching all the locations across multiple mini-maps, but the small size of the map glyphs themselves may also impact performance.

[H3] Identifying dates - *TimeG > CoordV & MapG*
As Time Glyphs group together all the relationships of one location, we expect it will perform better than the other two when identifying the dates when two people were linked with a location (`Baseline Time`).

[H4] Map Scalability - *CoordV > MapG > TimeG*
We expect that all of the designs will suffer as we increase the number of locations shown in the maps. But Time Glyphs' performance will likely deteriorate the most, as more locations means participants will have to visually aggregate information across more separate timelines. We also expect that Map Glyphs will suffer more than Coordinated Views due to the small size of the mini-maps. Increasing the number of locations on such small maps risks making them illegible.

5.4 Experiment Design and Procedure

We used a within-participant study design, where participants are exposed to all three visualizations. For each technique, they performed 10 training trials (2 repetitions per task), and 5 measured tasks × 4 repetitions = 20 main trials. In total, the experiment consists of 18 participants × 3 visualizations × 20 main trials = 1080 main trials.

We counterbalanced technique and dataset presentation order using a Latin square. Additionally, we rotated the maps used in the trials by 0, 90 and 180 degrees across trials, to avoid participants learning the location patterns on the maps. Finally, the task presentation order, was randomly assigned for each participant.

The study was divided into three 30 min sessions, one per visualization. At the start of the first session, participants signed a consent form and completed a demographics questionnaire. The rest of the sessions have the same structure: (i) explanation, (ii) training and (iii) main trials. In the explanation phase, an experimenter described the visualization, and guided participants in exploring the available interactions. The training phase consists of ten trials, two per task. Participants are asked to think aloud while completing the trials in order for the experimenter to provide clarifications when needed. If participants failed to complete a training trial, or ignored tools available to them (e.g., selection interactions) the experimenter would provide help. Participants only continued with the main trials if they completed all the training ones.

In the main trials, participants carried out four repetitions for each of the five tasks, using different datasets than those in training. In this phase, participants solved the tasks without instruction and did not follow a think aloud protocol. They were asked to perform the tasks as accurately and quickly as possible. After each trial, participants were asked to rate how easy it was to solve the task and how confident they were about their answer. At the end of each session, participants were asked to explain their strategies. And at the end of all three sessions, they were asked to rank the three visualizations as a whole.

5.5 Datasets

To avoid learning effects between visualizations and tasks, we created a separate dataset for each trial: 10 for training (the same for all visualizations); and 60 for the main trials (20 main trials × 3 variations, assigned to visualizations using a Latin Square). To ensure realistic temporal relationship patterns, our datasets are extracted from real world data provided by the news agency OuestFrance [4]. When entities of different types (people, locations) appear together in an article, this co-occurrence represents a relationship [41].

To make the study tractable our trials focused on data for two central people, adding another eight secondary people that were connected to them as distractors (*i.e.* a total of 10 people per dataset). This matches the use-case from data journalism described in the introduction, where

we can assume one has filtered their dataset to focus and follow two rival politicians. This total number of people allowed us to compare the basic properties of the designs, without considering the impact of searching or filtering functionality. Similar to previous studies [12, 39], we consider 35 time steps with one or more relationships per date. On average, an extracted dataset has 36 relationships ($SD = 1$), 10 of these are between the two central people (*i.e.*, 28% of the dataset on average).

We manually generated two spatial patterns for each of the five tasks. Each of these patterns were later projected into two maps of different size (one with 23 locations and the other with 59). This prevented participants from learning the answer for each task, and allowed us to test the geographic scalability of the designs. We replaced the original locations that dealt with regions in France with locations in maps unlikely to be known by participants, as did Beecham *et al.* [10].

Details of the dataset construction are available in supp. material. They include the scripts to generate the datasets, the maps, the spatial patterns and solutions, as well as steps taken to ensure trials were visually different but of equivalent difficulty across visualizations.

5.6 Participants & Apparatus

We recruited 18 participants, from 24 to 45 years old (*mean* = 28, $SD = 4.8$). None of them had participated in our design workshops. Five identified themselves as women and 13 as men. Seven of them had a background in HCI, six in visualisation, and five in other areas that included geology, data science or quantum computing. We aimed for a mix of participant backgrounds in order to elicit different types of feedback: participants with an HCI background can identify general usability issues, participants with a Vis background can provide a general critique of visual analytics systems, and participants with broader backgrounds represented a more general user population. 17 participants had normal, or corrected-to-normal vision. One reported color vision deficiency with the red component of colors, however, our color selection was such that they could conduct the study normally. Participants did not receive any monetary compensation.

We conducted the experiment on a Macbook Pro with a 27" Apple Thunderbolt Display (2560x1440 pixels). The system and visualizations were implemented in the Django framework [3] and Javascript respectively. The visualizations additionally use D3.js [2], and a publicly available Storyline layout algorithm [1].

5.7 Measures

We collect four primary measures per trial:
- `Completion time`: time in milliseconds from the moment a participant is presented with a task to when they select an answer.
- `Error rate`: percentage of incorrect answers per task.
- `Self-reported confidence`: a 5-point Likert scale (*Highly not confident*, *Not confident*, *Neutral*, *Confident*, and *Highly confident*).
- `Self-reported easiness to do the task`: as a 5-point Likert scale (*Very hard*, *Hard*, *Neutral*, *Easy*, *Very easy*).

Additionally, we collected participants' strategies per task, their overall feedback and preference of the three designs.

6 RESULTS OF COMPARATIVE STUDY

We report and interpret our results of Completion Time and Error Rate using interval estimation [16, 19]. Sample means of 95% confidence intervals (CIs) are constructed using BCa bootstrapping (10,000 bootstrap iterations). No p-values are reported, but they can be obtained from CI results [29]. When interpreting results, a CI that does not overlap with 0 provides evidence of a difference, corresponding to statistically significant results in traditional p-value tests. Nonetheless, CIs allow for more subtle interpretations: the farther from 0 and the tighter the CI is, the stronger the evidence. For self-reported confidence and easiness to complete the tasks, we only report means. All analyses were planned before collecting the data, and the study was preregistered [15] with the Open Science Framework[2]. We removed one trial from our analysis due to a technical problem. Analysis scripts, collected data (quantitative and qualitative), and detailed CIs can be found in supp.material.
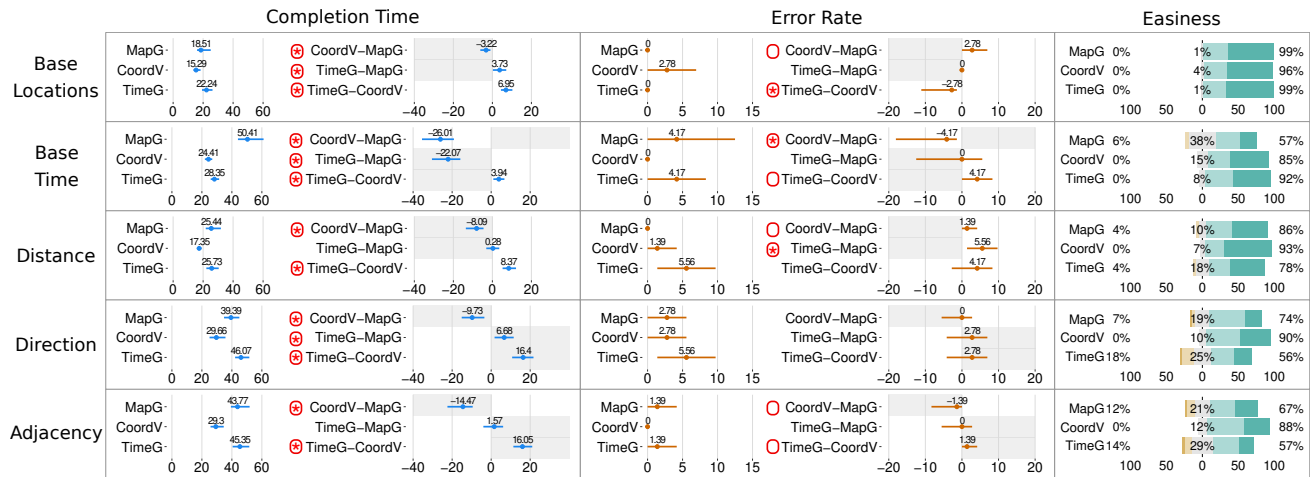
---

[2] `https://osf.io/5wnyg`

Fig. 3: Completion Time (sec), Error Rate (in %) and mean self-reported easiness per visualization per task. In each row (task) for the first two measures, mean values per visualization are seen on the left and means of pairwise differences on the right. Error bars represent 95% Bootstrap confidence intervals. Gray rectangles indicate the direction of our hypotheses. Evidence of differences is marked with a ⊛ (the further away from 0 and the tighter the CI, the stronger the evidence). Inconclusive evidence (barely touching 0) is not reported in text, but marked with ○ as it may represent trends. The 3rd column shows the percentage of trials that participants reported being Very easy (■) to complete, to very hard (■).

6.1 Overall performance for tasks collectively

For all tasks collectively, mean times are lower for Coordinated Views (23.2s), than for Time Glyphs (33.55s) and Map Glyphs (35.55s). There is evidence of Coordinated Views being faster than both Map Glyphs (by 12.35s on avg. CI[8.77:17.75]) and Time Glyphs (by 10.34s on avg. CI[7.92:12.66]). Although mean Error Rate is also lower for Coordinated Views (1.39%), followed by Map Glyphs (1.67%) and Time Glyphs (3.33%), there is no evidence of difference between the techniques. Self-reported confidence is high for the three techniques in more than 90% of the trials. Self-reported easiness to complete the task is high for 90% of the trials when using Coordinated Views, and 76% of the trials for Time Glyphs and Map Glyphs. The overall ranking was consistent with performance results: 15/18 participants ranked Coordinated Views as being best overall, followed by 5/18 raking Map Glyphs (participants were allowed to rank more than one technique as best). No participant ranked Time Glyphs as best. The results dis-aggregated by task (next) give more insights of these differences. From now on we do not report self-reported confidence as it is high for more than 80% of all trials, for all tasks and visualizations (see supp.material).

6.2 Results per working hypothesis

We analyze results for each task individually and group them by our initial hypotheses, displayed in gray next to them. Fig. 3 shows an overview of our results per task and technique. We do no report inconclusive evidence in the text (CIs that barely cross 0), but we highlight them in our image as it could indicate interesting trends for future study.

[H1] Movement patterns - *CoordV > MapG > TimeG*
For both tasks that analyzed movement, Direction and Adjacency, there is evidence that Coordinated Views are faster than both Map Glyphs and Time Glyphs. Additionally, there is evidence that Map Glyphs are faster than Time Glyphs in the Direction task (consistent with our hypothesis). There is no evidence of difference for Error Rate for these two tasks. Self-reported easiness to complete the task using Time Glyphs was lower than with the other two designs, while easiness was consistently higher with Coordinated Views.

There is support for **H1** as Coordinated Views are faster than the other two designs without being more prone to errors, and Map Glyphs are faster than Time Glyphs in one task. Participants mentioned that Map Glyphs required a lot of scrolling to explore the whole timeline, explaining why they were slower than Coordinated Views. Additionally, the gaps between the mini maps on the storyline required participants to rely on their memory of already visited mini maps in order to follow a movement pattern (10/18 participants commented on this issue). Time

Glyphs required participants to select and unselect dates in chronological order to see the evolution of the dynamic pattern across all storylines. This involved a lot of selection actions and participants had to rely on their memory, as they would continuously switch focus between the map and dates (11/18 participants commented about this).

[H2] Identifying locations - *CoordV & TimeG > MapG*
There is evidence that Coordinated Views are faster than Map Glyphs and Time Glyphs for tasks that require users to observe geographical entities (Baseline Locations and Distance). There is also evidence that Map Glyphs are faster than Time Glyphs in the Baseline Locations task. For Error Rate, Time Glyphs are less prone to errors than Coordinated Views for the Baseline Locations task, and Map Glyphs are less prone to errors than Time Glyphs in the Distance task.

We have partial support for **H2**. Coordinated Views are the fastest, but Time Glyphs are more accurate in Baseline Locations. What is surprising is that Map Glyphs were not the worst: they were either faster or less error prone than Time Glyphs in these tasks, and were even more accurate than Coordinated Views in Baseline Locations. We were surprised that participants were 100% accurate only with Map Glyphs. In our hypothesis we assumed that a single big map (Coordinated Views and Time Glyphs) would help participants to identify and compare locations more easily than several small ones (Map Glyphs).

[H3] Identifying dates - *TimeG > CoordV & MapG*
There is strong evidence that Map Glyphs are slower than both Coordinated Views and Time Glyphs for the task of Baseline Time. There is also evidence that Time Glyphs are slower than Coordinated Views. Regarding Error Rate, there is evidence that Coordinated Views were less prone to errors than Map Glyphs. Self-reported easiness to complete the task for Map Glyphs was high for a lower number of trials than the other two (only 57%).

There is no support for **H3** as Coordinated Views performed best in this task. Our hypothesis was based on the fact that Time Glyphs group all relationships for a location in one timeline. Therefore, the best strategy was to first select the location of interest and then the two people. However, only 2 participants used the optimized strategy consistently during the trials, 5 only used it during some of the trials and 8 used it only during training. We believe this result is due the fact that our tasks were designed to mainly focus on the behavior of two people, and not on particular locations. Thus our participants adopted a global strategy that was ineffective for this task.

[H4] Map Scalability - *CoordV > MapG > TimeG*
When looking at how map size affected each technique, we found no evidence of a differences for Map Glyphs. For Time Glyphs there

seems to be a time/accuracy trade-off: tasks on smaller maps led to more errors but were faster. Most surprising, in Coordinated Views performance using small maps was worst both for time and errors. We thus did not find support for **H4**, as it seems that smaller maps either were not easier or were in fact worst. For detailed CIs see supp.material.

7 FEEDBACK FROM DOMAIN EXPERTS

We complemented the results of our user study with feedback from domain experts that analyse geo-temporal relationships. We conducted two workshops in the context of a long term collaboration with Ouest France [4], the most read francophone newspaper in the world (see work on HyperStorylines [41]). In a first informal workshop (7 investigative journalism practitioners) that lasted 1h, experts expressed their interest in combining storylines with maps in order to analyse the movement and relationships of politicians. They even suggested the generalization of the concept by linking storylines with other 2D visualizations (for example a thematic map instead of a geographical one).

We then conducted a structured 1.5h workshop with three investigative journalism practitioners (more than 10 years of experience each). We asked our domain experts to explore their own data using the three final Geo-Storylines designs. The dataset they provided contained a set of news articles about a famous local politician who, over time, changed the region they represented in local elections. It contained 32 articles with 102 people entities over the span of 4 years. The map showed 1272 locations (communes). The dataset is not publicly available due to a confidentiality agreement.

We provided brief descriptions of each visualization and then let them explore their dataset. To help guide the exploration, we asked them (in each vis) if they can describe the overall movement of the protagonist politician. One participant (P1) interacted with the system and the other two (P2,3) suggested actions and provided comments.

In Coordinated Views, P1 was able to quickly identify the overall trajectory of the politician. P1 and P2 explained it was very helpful to have a mix of a storyline and a map, as it helped them to explore the dataset starting either with people or locations of interest. Even though they were able to find the overall movement pattern, they requested more explicit ways to see the patterns. For example, P1 suggested adding arrows on the map to show explicit movement, and a "play" button to provide automatic scrolling of the timeline. P3 requested the addition of a mechanism to extract several important stories, *e.g.,* extract the relationships between the top 1-2 most connected politicians.

Although Time Glyphs performed worse than Coordinated Views in our study, experts appreciated its ability to show stories per commune / location. P1 commented (and the others agreed): *"It's hard to see the [person's] story, but we can see clearly the articles for each location which is very important when analysing local news."* Both P1 and P2 requested to completely filter out (remove) the storylines of uninteresting locations. Finally, all experts mentioned how Time Glyphs and Coordinated Views are complementary, and both are useful depending on the focus of the exploration (politicians vs. local councils/politics).

Participants found Map Glyphs somewhat overwhelming. Given the large number of communes (more than 1200 locations), each location ended up being very small in the map glyphs, thus it was hard to see which communes were involved in several relationships. Our participants requested some form of geographic aggregation. P1 commented: *"It would be interesting to choose the zoom level [geographical granularity] of the visualization"*. After exploring some more, P1 and P2 mentioned that one article per relationship was too much information, as each relationship introduced a new map glyph in the visualization. They suggested to filter or aggregate relationships by time.

Participants were overall very enthusiastic about the exploration potential of the visualizations, and at the end of the session they requested an open link to the software to continue exploring their data.

8 DISCUSSION, LIMITATIONS & FUTURE WORK

We discuss next the results of our study and feedback from experts on the three most promising geo-storyline designs. Then we take a step back to consider the limitations of our study and design space.

8.1 Study Results and Expert Feedback

8.1.1 Identifying entities and overall navigation

Coordinated Views performed well when considering all study tasks combined, likely because it presents a more compact timeline than the other two designs. Issues mentioned by our study participants and journalism experts can start explaining these results.

Map Glyphs do not offer an overview of the data as the storyline expands to accommodate the map glyphs, so participants had to scroll a lot to do the tasks. Five study participants commented on this lack of overview for Map Glyphs (versus only two for Time Glyphs and none for Coordinated Views). Some mentioned specifically too much scrolling being an issue: eight for Map Glyphs, versus two for Coordinated Views and one for Time Glyphs. Our domain experts also commented on the clutter of the Map Glyphs design that makes it hard to follow the entity relationships. We can think of several ways to improve Map Glyphs in this regard. First, we can allow users to customize the size of the mini-maps, adjusting the importance the geographical context has in the view. However, it is possible that users would like to focus on a particular set of relationships so the importance of the mini-maps might not be the same across the timeline. Therefore, an alternative could be to make the relationships (and thus mini-maps) collapsible by default and only open on demand, in a way similar to nested entities in HyperStorylines [41]. Or, as some domain experts suggested, to aggregate relationships over time (over a week or a month) thus reducing the number of required mini-maps.

In our study, finding a particular person with Time Glyphs takes longer as people's names are scattered over the different timeline glyphs. Indeed, six participants had at least one comment about this drawback (versus two participants for Map Glyphs and only one for Coordinated Views). We believe it is hard to improve this design without applying general strategies like adding a list of entities or a search option. Nevertheless, we did note this design was less error-prone than Coordinated Views when it came to identifying and counting locations themselves. Contrary to the study results, our domain experts really appreciated that when using Time Glyphs they could focus on the story of specific locations (focusing on local news and politics). They explained that Time Glyphs and Coordinated Views are complementary designs.

We were surprised to see that Map Glyphs, which have small map representations, reported no errors in the tasks related to identifying locations in our study (Baseline Locations and Distance), and weree more accurate than the other two designs (including Coordinated Views). It is possible that seeing fewer details of the maps allowed participants to be less distracted by their geographic details. However map size was an issue for our experts with this design: as their dataset contained a very large number of locations (more than a thousand) the mini maps were too cluttered and locations too small to interact with.

8.1.2 Understanding Movement Patterns

Coordinated Views performed well in the two tasks related to movement patterns (Direction and Adjacency). As the horizontal scrolling simulates an animation of the geographical pattern, our results are aligned with our hypotheses and previous work that have shown that sudden changes are better detected with animation [14, 23, 39]. Our domain experts confirmed this, requesting automatic scrolling (play an animation) on demand of the entire timeline. Both study participants and our experts found navigation with Map Glyphs to be hard and tedious as the resulting timeline is bigger than the other two designs. In addition, the mini-maps are dispersed across the timeline, and the gaps between them makes their sequential comparison (to follow a pattern) harder. The collapsable or aggregated mini-maps mentioned earlier can improve this issue by bringing them closer for comparison. However, we hypothesize that it will remain less efficient than Coordinated Views.

Time Glyphs required participants to shift their focus across regions of the visualization to follow the entities of interest. The best strategy was to select the dates in chronological order to highlight the locations associated to each time step, switching focus between the dates and the map. This process can be improved by implementing an animation option in the interface that automatically highlights progressively the

dates and associated locations for the selected people. We hypothesize this would make Time Glyphs comparable to Coordinated Views for observing movement patterns, but this remains future work. We note nonetheless, that experts felt Time Glyphs excelled in cases where the focus is on local news (small number of locations). It is likely that our study tasks did not reflect well this real-world need.

8.1.3 Number of locations and map complexity

We expected that the number of locations would impact performance, in particular for Map Glyphs that use small maps to begin with. Surprisingly we only saw clear evidence of a difference between the two types of maps in Coordinated Views, where trials using the smaller map with fewer locations was in fact more prone to errors. While further experimentation is needed, we feel this may be due to the nature of the chosen maps: our small map may have been more visually complex, as it includes locations of very different sizes (including some very small ones) and water features that participants could have confused with locations. Further studies are needed where more diverse metrics of geographical complexity are used to measure this effect, *e.g.,* geometric irregularity [10, 25]. Feedback from our domain experts provided more insights on scalability, as they used the visualizations with maps of more than a 1000 locations. Our experts did not raise any issues with Coordinated Views. For Time Glyphs they communicated the need to filter out unused location storylines to reduce clutter. More importantly, they found location scalability to be an issue for Map Glyphs, as locations become too small to interact with. They suggested aggregating geographical locations, allowing semantic zooming to see content at different granularity (*e.g.,* region vs. country), thus reducing the number of locations on mini-maps, making them more readable.

8.2 Study: Discussion & Limitations

For our evaluation, we chose three representative designs that could easily accommodate several people, locations and temporal relationships, without the need to aggregate any of these entities. Even though our datasets had 35 timesteps (comparable to previous work [12, 39]), the number of people in our tasks (10) was smaller than that of previous work focusing on entity relationships (HyperStorylines [41]: around 529 entities, or PAOHVis [55]: 57-371 people). This was a deliberate decision, as it allowed us to compare the benefits and drawbacks of the basic designs, without adding search and filtering interactions (which were necessary in previous work with more entities [41, 55]).

Nevertheless, in real analysis situations our designs would have to handle larger numbers of people or locations, as our expert feedback demonstrated. We could envision only transitioning to the geo-spatial views after the data has been filtered, as it was suggested by our domain experts (*e.g.,* focusing on the most connected politicians). But more interesting yet would be to consider aggregation, to provide meaningful overviews: our experts suggested aggregating relationships over time to handle the number of relationships, and geographical aggregation of cities/regions/countries to handle large number of locations. It is interesting to note that our experts raised these scalabilty issues mainly for Map Glyphs. Further research is required to determine the best aggregation methods of entities evolving over time [42], how to allow viewers to determine the required level of detail on the fly, and how our design space can accommodate such aggregations.

Our study tasks focus on people whose relationships follow five specific spatial patterns. We removed the patterns Hierarchy and Coverage as the strategies adopted in our pilots were similar to our task `Baseline Locations`. However, if we consider explicit aggregations to deal with scale, it is likely that the strategies for some of these tasks may diverge (*e.g.,* Hierarchy). While our chosen tasks cover a wide range of patterns from the related work, it is possible that they may have biased against some of our techniques. For example, Time Glyphs could have been more efficient for questions that focus on the co-occurrence of locations. This is supported by feedback from our experts that found Time Glyphs very useful for focusing on news articles (relationships) from specific regions. More generally, it remains future work to investigate whether our results hold for patterns about other types of entities (not just people), as well as more complex movement patterns (more central

people, more relationships, etc.), especially since these would require filtering and aggregation options as discussed above.

Finally, 2/3 of the participants of our user study had an HCI or visualization background. These participants could be considered as experts in the use of interactive systems, and thus find the tools easier to use than the general public. So we complemented our results with feedback from domain experts in data journalism. Further work can explore the use of these designs in other domains or general public use.

8.3 Design space: Discussion & Limitations

When organizing the workshop we provided participants with inspiration visualizations that were not necessarily storyline or timeline centered. Nevertheless, a large number of participant sketches naturally converged to such representations, confirming the ability of storylines to convey temporal relationships. We did categorize all sketches from the workshop, but for our design space we kept ones that could accommodate multiple people and their relationships (often as timelines), but also provide geo-spatial context in the form of a map, rather than an abstract representation of space (like a chart indicating distance).

In our designs, maps are divided in regions. As discussed, this fixed level of granularity may not always be appropriate, as geo-spatial navigation could adapt the content depending on zoom level (semantic zooming). Future work should extend the design space to accommodate transitions between different levels of geo-spatial granularity. For example, in Time Glyphs, the storylines for countries can be split into storylines for cities as a response to zooming-in. Or in Map Glyphs and Coordinated Views, regions can be semantically merged or split to the level of cities, regions or countries as the user zooms out or in.

Our design space did not consider interactions and we implemented basic interaction support (scroll, selection). Nevertheless, the changes in granularity discussed above require interaction mechanisms, like semantic zooming, interactively choosing appropriate levels of aggregation per entity, and animated transitions. This, combined with user requests for filtering and switching views, points to the need for future design spaces and user studies that consider advanced interaction and animated transitions. This may require revisiting some of the discarded designs of the design space that rely on aggregation.

Our aim was to provide geo-spatial context into storylines, which naturally limits our designs to representations of "space" that are geographical maps. But our design space can be extended to combine storylines with other 2D visualizations that represent space or different dimensions. Consider for example scatterplots or treemaps. As long as the definition of a region is possible (a treemap cell, a scatterplot selection or cluster), then our designs can combine these 2D visualizations with storylines. For example, we can envision customizable timelines that are created after a manual selection over a scatterplot similarly to Attribute Signatures [54], resembling our Time Glyphs design but focusing on the selected area. Or instead of a geographical map, Coordinated Views and Map Glyphs may show a treemap, highlighting the relevant cells. Nevertheless, if the number of 2D regions is large, or the regions themselves are small, some designs from our design space will suffer: for example it will be challenging to create one storyline per point in a scatterplot (Time Glyphs), or highlight individual scatterplot points in a small 2D glyph embedded in a large storyline (Map Glyphs).

9 CONCLUSIONS

We presented Geo-Storylines, a set of designs to integrate maps into Storyline visualizations. We report on a sketching workshop used to derive a design space for their creation, the arguments to select the three more promising designs, and the results of a user study and expert feedback that compares them. Our study shows that, overall, Coordinated Views were more efficient and preferred than the other two designs. Surprisingly, although Map Glyphs reported no errors in the user study, they were hard to use with a real-world dataset. Finally, even though Time Glyphs did not outperform other techniques in our user study when focusing on a particular location, they were highly appreciated by our expert users. Our work illustrates how Storyline visualizations can be effectively extended to incorporate 2D information, including maps, improving their already considerable expressive power.

REFERENCES

[1] D3 layout narrative. `https://github.com/abcnews/d3-layout-narrative`. Last accessed: 31.03.2021.

[2] D3.js. `https://d3js.org/`. Last accessed: 31.03.2021.

[3] Django. `https://www.djangoproject.com/`. Last accessed: 31.03.2021.

[4] Ouest france. Accessed on 31.03.2021.

[5] Star wars: every scene from i-vi charted. `https://www.abc.net.au/news/2015-12-16/star-wars-every-scene/7013826`. Last accessed: 01.07.2022.

[6] N. Andrienko and G. Andrienko. *Exploratory analysis of spatial and temporal data: a systematic approach*. Springer Science & Business Media, 2006.

[7] D. Arendt and M. Pirrung. The "y" of it matters, even for storyline visualization. In *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 81–91, 2017. doi: 10.1109/VAST.2017.8585487

[8] B. Bach, P. Dragicevic, D. Archambault, C. Hurter, and S. Carpendale. A descriptive framework for temporal data visualizations based on generalized space-time cubes. *Computer Graphics Forum*, 36(6):36–61, 2017. doi: 10.1111/cgf.12804

[9] J. T. Balint, D. Arendt, and L. M. Blaha. Storyline visualizations of eye tracking of movie viewing. In *2016 IEEE Second Workshop on Eye Tracking and Visualization (ETVIS)*, pp. 35–39, 2016. doi: 10.1109/ETVIS.2016.7851163

[10] R. Beecham, J. Dykes, W. Meulemans, A. Slingsby, C. Turkay, and J. Wood. Map LineUps: Effects of spatial structure on graphical inference. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):391–400, Jan. 2017. doi: 10.1109/TVCG.2016.2598862

[11] C. Berge. *Hypergraphs: combinatorics of finite sets*, vol. 45. Elsevier, 1984.

[12] I. Boyandin, E. Bertini, and D. Lalanne. A qualitative study on the exploration of temporal changes in flow maps with animation and small-multiples. *Computer Graphics Forum*, 31(3pt2):1005–1014, 2012. doi: 10.1111/j.1467-8659.2012.03093.x

[13] M. Brehmer, B. Lee, B. Bach, N. H. Riche, and T. Munzner. Timelines revisited: A design space and considerations for expressive storytelling. *IEEE Transactions on Visualization and Computer Graphics*, 23(9):2151–2164, 2017. doi: 10.1109/TVCG.2016.2614803

[14] M. Brehmer, B. Lee, P. Isenberg, and E. K. Choe. A comparative evaluation of animation and small multiples for trend visualization on mobile phones. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):364–374, 2019.

[15] A. Cockburn, C. Gutwin, and A. Dix. Hark no more: On the preregistration of chi experiments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pp. 141:1–141:12. ACM, 2018. doi: 10.1145/3173574.3173715

[16] G. Cumming. The new statistics: Why and how. *Psychological Science*, 25(1):7–29, 2014. doi: 10.1177/0956797613504966

[17] E. Di Giacomo, W. Didimo, G. Liotta, F. Montecchiani, and A. Tappini. Storyline visualizations with ubiquitous actors. In *Graph Drawing and Network Visualization*, Lecture Notes in Computer Science, pp. 324–332. Springer International Publishing, 2021.

[18] S. Dodge, R. Weibel, and A.-K. Lautenschütz. Towards a taxonomy of movement patterns. *Information Visualization*, 7(3-4):240–252, 2008. doi: 10.1057/PALGRAVE.IVS.9500182

[19] P. Dragicevic. Fair statistical communication in HCI. In J. Robertson and M. Kaptein, eds., *Modern Statistical Methods for HCI*, chap. 13, pp. 291–330. Springer, Cham, Switzerland, 2016.

[20] J. Eisenstein, B. T. O'Connor, N. A. Smith, and E. P. Xing. Diffusion of lexical change in social media. In *PloS one*, 2014. doi: 10.1371/journal.pone.0113114

[21] M. T. Fischer, D. Arya, D. Streeb, D. Seebacher, D. A. Keim, and M. Worring. Visual analytics for temporal hypergraph model exploration. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):550–560, 2021. doi: 10.1109/TVCG.2020.3030408

[22] M. T. Fischer, A. Frings, D. A. Keim, and D. Seebacher. Towards a survey on static and dynamic hypergraph visualizations, 2021.

[23] A. L. Griffin, A. M. MacEachren, F. Hardisty, E. Steiner, and B. Li. A comparison of animated maps with static small-multiple maps for visually identifying space-time clusters. *Annals of the Association of American Geographers*, 96(4):740–753, 2006. doi: 10.1111/j.1467-8306.2006.00514.x

[24] M. Gronemann, M. Jünger, F. Liers, and F. Mambelli. Crossing minimization in storyline visualization. In *Graph Drawing and Network Visualization*, Lecture Notes in Computer Science, pp. 367–381. Springer International Publishing, 2016.

[25] J. D. Haider, M. Pohl, R. Beecham, and J. Dykes. Strategies for detecting difference in map line-up tasks. In C. Ardito, R. Lanzilotti, A. Malizia, H. Petrie, A. Piccinno, G. Desolda, and K. Inkpen, eds., *Human-Computer Interaction – INTERACT 2021*, pp. 558–578. Springer International Publishing, Cham, 2021.

[26] S. Huron, P. Gourlet, U. Hinrichs, T. Hogan, and Y. Jansen. Let's get physical: Promoting data physicalization in workshop formats. In *Proceedings of the 2017 Conference on Designing Interactive Systems*, DIS '17, p. 1409–1422. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3064663.3064798

[27] M. Itoh, D. Yokoyama, M. Toyoda, Y. Tomita, S. Kawamura, and M. Kitsuregawa. Visual exploration of changes in passenger flows and tweets on mega-city metro network. *IEEE Transactions on Big Data*, 2(1):85–99, 2016. doi: 10.1109/TBDATA.2016.2546301

[28] N. W. Kim, S. K. Card, and J. Heer. Tracing genealogical data with timenets. In *Proceedings of the International Conference on Advanced Visual Interfaces*, AVI '10, p. 241–248. Association for Computing Machinery, New York, NY, USA, 2010. doi: 10.1145/1842993.1843035

[29] M. Krzywinski and N. Altman. Points of significance: Error bars. *Nature Methods*, 10(10):921–922, 10 2013. doi: 10.1038/nmeth.2659

[30] F. Lekschas, X. Zhou, W. Chen, N. Gehlenborg, B. Bach, and H. Pfister. A generic framework and library for exploration of small multiples through interactive piling. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):358–368, 2021. doi: 10.1109/TVCG.2020.3028948

[31] S. Liu, Y. Wu, E. Wei, M. Liu, and Y. Liu. Storyflow: Tracking the evolution of stories. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2436–2445, 2013. doi: 10.1109/TVCG.2013.196

[32] M.-J. Lobo, C. Appert, and E. Pietriga. Animation plans for before-and-after satellite images. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 25(2):1347–1360, Feb 2019. doi: 10.1109/TVCG.2018.2796557

[33] Q. Lu, X.-y. Zhu, L. Liu, and S.-b. Cao. An effective demonstration for group collaboration based on storyline visualization technology. In *Proceedings of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pp. 47–52, 2014. doi: 10.1109/CSCWD.2014.6846815

[34] R. Maciejewski, P. Livengood, S. Rudolph, T. F. Collins, D. S. Ebert, R. T. Brigantic, C. D. Corley, G. A. Muller, and S. W. Sanders. A pandemic influenza modeling and visualization tool. *Journal of Visual Languages & Computing*, 22(4):268 – 278, 2011. doi: 10.1016/j.jvlc.2011.04.002

[35] R. Munroe. Movie narrative charts, 2009. Accessed on 28.06.2021.

[36] M. Ogawa and K.-L. Ma. Software evolution storylines. In *Proceedings of the 5th International Symposium on Software Visualization*, SOFTVIS '10, p. 35–42. Association for Computing Machinery, New York, NY, USA, 2010. doi: 10.1145/1879211.1879219

[37] K. Padia, K. H. Bandara, and C. G. Healey. Yarn: Generating storyline visualizations using htn planning. In *Proceedings of the 44th Graphics Interface Conference*, GI '18, p. 26–33. Canadian Human-Computer Communications Society, Waterloo, CAN, 2018. doi: 10.20380/GI2018.05

[38] K. Padia, K. H. Bandara, and C. G. Healey. A system for generating storyline visualizations using hierarchical task network planning. *Computers & Graphics*, 78:64–75, 2019. doi: 10.1016/j.cag.2018.11.004

[39] V. Peña Araya, A. Bezerianos, and E. Pietriga. A comparison of geographical propagation visualizations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–14. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3313831.3376350

[40] V. Peña-Araya, E. Pietriga, and A. Bezerianos. A Comparison of Visualizations for Identifying Correlation over Space and Time. *IEEE Transactions on Visualization and Computer Graphics*, Oct. 2019. doi: 10.1109/TVCG.2019.2934807

[41] V. Peña-Araya, T. Xue, E. Pietriga, L. Amsaleg, and A. Bezerianos. Hyperstorylines: Interactively untangling dynamic hypergraphs. *Information Visualization*, 0(0):14738716211045007, 0. doi: 10.1177/14738716211045007

[42] A. Pister, P. Buono, J.-D. Fekete, C. Plaisant, and P. Valdivia. Integrating prior knowledge in mixed-initiative social network clustering. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1775–1785,

2021. doi: 10.1109/TVCG.2020.3030347

[43] L. Qiang, C. Bingjie, and Z. Haibo. Storytelling by the storycake visualization. *The Visual Computer*, 33(10):1241–1252, 2017. doi: 10.1007/s00371-017-1409-2

[44] K. Reda, C. Tantipathananandh, A. Johnson, J. Leigh, and T. Berger-Wolf. Visualizing the evolution of community structures in dynamic social networks. *Computer Graphics Forum*, 30(3):1061 – 1070, 2011.

[45] H. Romat, D. Lebout, E. Pietriga, and C. Appert. Influence of color and size of particles on their perceived speed in node-link diagrams. In *Human-Computer Interaction – INTERACT 2019*, pp. 619–637. Springer International Publishing, Cham, 2019. doi: 10.1007/978-3-030-29384-0_37

[46] S. Schöttler, Y. Yang, H. Pfister, and B. Bach. Visualizing and interacting with geospatial networks: A survey and design space. In *Computer Graphics Forum*. Wiley Online Library, 2021.

[47] S. Schöttler, Y. Yang, H. Pfister, and B. Bach. Visualizing and interacting with geospatial networks: A survey and design space. *Computer Graphics Forum*, 40(6):5–33, 2021. doi: 10.1111/cgf.14198

[48] Y. Tanahashi, C.-H. Hsueh, and K.-L. Ma. An efficient framework for generating storyline visualizations from streaming data. *IEEE Transactions on Visualization and Computer Graphics*, 21(6):730–742, 2015. doi: 10.1109/TVCG.2015.2392771

[49] Y. Tanahashi and K.-L. Ma. Design considerations for optimizing storyline visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2679–2688, 2012. doi: 10.1109/TVCG.2012.212

[50] T. Tang, R. Li, X. Wu, S. Liu, J. Knittel, S. Koch, T. Ertl, L. Yu, P. Ren, and Y. Wu. Plotthread: Creating expressive storyline visualizations using reinforcement learning. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):294–303, 2021. doi: 10.1109/TVCG.2020.3030467

[51] T. Tang, S. Rubab, J. Lai, W. Cui, L. Yu, and Y. Wu. istoryline: Effective convergence to hand-drawn storylines. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):769–778, 2019. doi: 10.1109/TVCG.2018.2864899

[52] S. Thakur and A. J. Hanson. A 3d visualization of multiple time series on maps. In *2010 14th International Conference Information Visualisation*, pp. 336–343, 2010. doi: 10.1109/IV.2010.54

[53] X. Tricoche, W. Schlei, and K. C. Howell. Extraction and visualization of poincare map topology for spacecraft trajectory design. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):765–774, 2021. doi: 10.1109/TVCG.2020.3030402

[54] C. Turkay, A. Slingsby, H. Hauser, J. Wood, and J. Dykes. Attribute signatures: Dynamic visual summaries for analyzing multivariate geographical data. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2033–2042, 2014. doi: 10.1109/TVCG.2014.2346265

[55] P. Valdivia, P. Buono, C. Plaisant, N. Dufournaud, and J.-D. Fekete. Analyzing dynamic hypergraphs with parallel aggregated ordered hypergraph visualization. *IEEE Transactions on Visualization and Computer Graphics*, 27(1):1–13, 2021. doi: 10.1109/TVCG.2019.2933196

[56] T. von Landesberger, F. Brodkorb, P. Roskosch, N. Andrienko, G. Andrienko, and A. Kerren. Mobilitygraphs: Visual analysis of mass mobility dynamics via spatio-temporal graphs and clustering. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):11–20, 2016. doi: 10.1109/TVCG.2015.2468111

[57] F. Windhager, P. Federico, S. Salisu, M. Schlögl, and E. Mayr. A synoptic visualization framework for the multi-perspective study of biography and prosopography data. In *Proceedings of the 2nd IEEE VIS Workshop on Visualization for the Digital Humanities (VIS4DH'17), Phoenix, AZ, USA*, vol. 2, p. 2017, 2017.

[58] S. Yagi, T. Itoh, and M. Takatsuka. A layout technique for storyline-based visualization of consecutive numerical time-varying data. In *Proceedings of the 8th International Symposium on Visual Information Communication and Interaction*, VINCI '15, p. 156–157. Association for Computing Machinery, New York, NY, USA, 2015. doi: 10.1145/2801040.2801067

[59] J. Zhao, M. Glueck, F. Chevalier, Y. Wu, and A. Khan. *Egocentric Analysis of Dynamic Networks with EgoLines*, p. 5003–5014. Association for Computing Machinery, New York, NY, USA, 2016.