

2 Making sense of spatial big data

Technology is the answer, but what was the question?

Cedric Price

This chapter first outlines the types of point spatial data, then explores the methods of spatio-temporal knowledge discovery. Then we explore how cartography can support understanding the world through the lens of big data. In conclusion, some objections to using visualisation to generate insight are discussed.

2.1 Spatial big data classification: stations, events, and agents

The vast majority of what is presently understood as spatial big data has point spatial reference. This prevalence comes naturally if we realize that the “data point” location is described basically as a coordinate pair – two digits that can be easily stored in standard database systems without the need to observe topological rules and other constraints that GIS vector data model enforces on line and polygon geometries. Point data are spatial data that are easily created and handled by non-spatial (meaning not GIS-enabled) systems that account for majority of data production. For this reason and due to the scope limits of this thesis, we will almost exclusively focus on visualisation issues related to point data¹.

Point spatial data are not a homogeneous group. We can describe three kinds of objects differentiated by their behaviour in space and time. More precisely, the difference

¹ We'll use the term *point data* as a shorthand for “data with point spatial reference”.

is in how dynamic the object's *existence*, *location* and *attributes* are over the course of observation. These properties are determined by the source of data, so for convenience we can label the three object types as *stations*, *agents* and *events*:

- stationary objects (*stations*) have static position and existence, meaning that they don't move or disappear during observation. What is dynamic is the set of attributes attached to the object – in big data world these attributes can come as a continuously updated streams. Basic examples include weather stations, traffic cameras, or any kind of stationary sensors.
- moving objects (*agents*) move around, so their position changes during observation, also their existence can be dynamic, meaning they can enter or exit the area of interest. Various kinds of dynamic attributes can be attached. We can reconstruct the history of movement of these objects, which invites conversion to linear representation. Examples are vehicles or pedestrians carrying GPS and sensor equipped devices.
- episodic objects (*events*) exist in a specific point of space and time. As they are short-lived, we can say that position and associated attributes are static. Prime example are data collected from social networks (tweets, posts, comments, etc.)

The difference between stations and events is dependent on the frame of reference, as objects seen as stationary in shorter observation periods can become mere events if the observation time frame is significantly extended. For example, the existence of a building usually spans over decades, though if we stretch the perspective to a century or a millennium, most buildings will become glimpses existing a tiny fraction of time². Geographers would note that also

² see <https://waitbutwhy.com/2013/08/putting-time-in-perspective.html> for a visualisation of perspectives changing with time frame

the location of seemingly static environmental features doesn't hold over time (think of a meandering riverbed or a volcanic landscape). So again, longer time frame changes our assumptions of static location.

Furthermore, the spatial extent of the observed area and hence the scale of the map influences the distinction between moving and stationary objects – if the movement is too limited to be recognized at a given scale, we can model it as a stationary object. Also, some events can be reimagined as moving objects with discrete presence across observation time frame, for example if social media events dislocated in space and time are traced back to a single moving source device³.

These notes underline that the distinction to stations, agents and events is just a convenience model that works because most big data sources are temporally and spatially limited to near real time and urban environment⁴. Judging by the real data samples we can say that stations are usually physically present in the environment while events are mainly records of something that happened at the location, either physically expressed in the environment and observable by onlookers (“I was at a restaurant”) or not (“I was shopping online while waiting at the bus stop”).

³ See examples of such practice at <https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html>

⁴ There are notable global-scale exceptions like <https://globalfishingwatch.org/map/?locale=en> or <https://www.shipmap.org/>

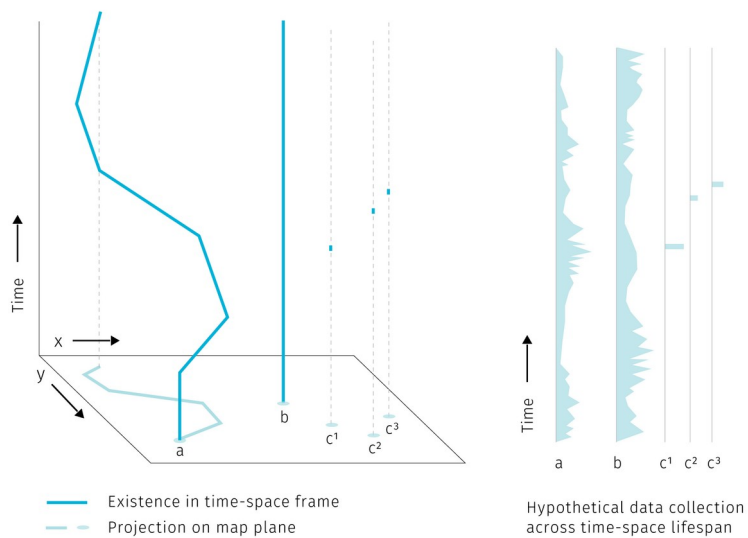


Fig. Three types of point spatial objects in a time-space cube. Stations, actors and events generate different attribute histories.

Tab1 Properties of point spatial object. Existence is marked by records of spatial and temporal reference. Agent can have discrete existence if exiting and re-entering the area of interest.

type of object	existence	attribute collection	location
continuous or discrete	static	station	continuous
continuous or discrete	dynamic	agent	continuous or discrete
discrete	static	event	discrete

In the above image and a table we assume that the attribute collection is happening continuously for stations and agents. This does not mean that the attributes have to be collected continuously at all times. Some sensors can record at a regular time interval or only in case of an event. The data output can then contain several “no data” records or even no

records at all if the triggering event did not happen. It then depends on the goal of the analysis how such data are conceptualized. For example a traffic camera is a stationary object but some part of its data collection is episodic – a photo is taken just when a speeding vehicle drives by. The above classification differentiates between the existence of an object and the act of recording data by the object. We assume that the sensor's presence without recording has also some analytical potential as it proves the absence of event, while with no sensor in place we cannot say if the event did take place or not.

Compared to stations and agents, events with episodic presence seem to be the least data-rich, but their analytic potential stems from their large numbers. Clusters of georeferenced point events, a.k.a. point clouds are at the core of spatial analysis based on mobile data.

2.2 Spatio-temporal knowledge discovery and visual analytics

In this section we will briefly discuss techniques for exploring spatio-temporal data, with emphasis on practices that would benefit from enhanced cartographic visualisation.

The expectation that motivates people engaged in data-related practices is that their work can help to provide some insight into how the world works, that there is some knowledge that can be unlocked, mined, or distilled from otherwise inconceivable piles of data. Such insight seeking is the crux of *data mining*, *spatio-temporal knowledge discovery* and *visual analytics* that we will explore further.

Data mining is exploring databases using low-level algorithms to find patterns. *Knowledge discovery* is then a higher-level extension of data-mining techniques that

requires human-level intelligence and domain knowledge to guide the process and interpret the results (Miller (2015)). In the knowledge discovery process, computation is seen as an extension of human force rather than its replacement — the goal is to marry the best of the both worlds. This is in line with the (current) capabilities of information technologies: there are tasks that are very simple for computers and very hard for humans (e.g. calculate the square root of 567789898) and vice-versa (basically any task requiring imagination and improvisation).

If we imagine a continuum ranging from “work done purely in human brain” towards “work done by machines”, knowledge discovery places itself somewhere in the middle. *Visual analytics*, the science of analytical reasoning supported by interactive visual interfaces (Thomas & Cook (2005)), then zooms in at the human-machine frontier in order to find the best tools for *visual* interaction.



Fig. Human-machine continuum, knowledge discovery as the best from the both worlds (the wording could be different, for example Keim et al. (2008) lists on the “machine” side: statistical analysis, data management, data mining, compression and filtering; on the “human” side: cognition, perception, visual intelligence, decision making theory, information design; and in the “middle”: human-centered computing, semantics-base approaches, graphics and rendering, and information visualisation). With emphasis on cartography, we summarize the human cognitive tasks as “map reading”.

We can very well imagine the human-machine continuum in the field of digital cartography. Here, the human cognitive abilities are applied to seek patterns, explore spatial context

or to make decisions, while computational aspects include data management and processing. The computation heavy algorithms like optimal route calculation already step in to unburden people from some decision making so the distinction shouldn't be taken as something rigid. Cartography provides an interface at the human side. Some authors go on to define *visual analytics for spatio-temporal data* as interlinked techniques in interfaces with map as a central metaphor (Guo, Chen, MacEachren, & Liao (2006)). We can think of it as map reading with robot assistants.

2.2.1 Spatio-temporal relations

To develop further on the kinds of interaction with spatial data, we can explore the concept of *spatial* and *temporal* queries. On the general level we can search for spatial and temporal relations in all there types of point objects mentioned in the first section. In addition, moving agents can generate specific relations not innate to stations and events.

Spatial relations are at the very basis of map reading for orientation clues, but are also vital for interpreting thematic information. We perceive these relations between the dominant themes (e.g. in weather maps of precipitation and atmospheric pressure zones) or between the theme and the topographical base map. The major classes of spatial relations are: *set-oriented* (union, difference, intersection, complement, etc.), *topological* (connectivity, interior, exterior, boundary), *directional* (cardinal, object-centered, ego-centered directions) and *metric* (e.g. Euclidean or network-based distance) (Worboys & Duckham (2004)).

Point spatial data of large extent complicate observing such relations. We rarely ask about a single specific point from the set, more often we seek to extract some tendency of the whole point cloud. The nature of some data sources can

dictate some spatial relationships (such as vehicles being spatially bound the road network), but in many cases the density of point cloud obscures the base map and precludes depicting of attribute variability within the set.

Spatial relations between point clusters are harder to conceptualize than it is with polygonal features. Egenhofer & Franzosa (1991) describe 16 types of spatial relations (9 if reduced to spatial regions relevant in GIS) in two dimensional space. However, in their approach Egenhofer & Franzosa (1991) define the point sets by their exterior boundary and then treat them as polygons. But delineating the exterior boundary is a challenge in itself, for example when dealing with smooth transitions in point density at the border, or when outliers are present. Spatial relations between point clouds in three dimensions are the subject of extensive research in the fields of computer vision and indoor navigation (e.g tran2017extracting or chen2019deep). However, the motivation here is object identification. In these lines of research the point cloud is representing distinct solid objects in the real space that need to be extracted, so the point cloud itself is not an object of research. For cartography, the point sets already come with some assigned attributes, so there is usually no need to label them algorithmically. Large point sets tend to get unruly, and saying anything meaningful about spatio-temporal relations of multiple such clouds is increasingly challenging for the basic set theory (see Fig).

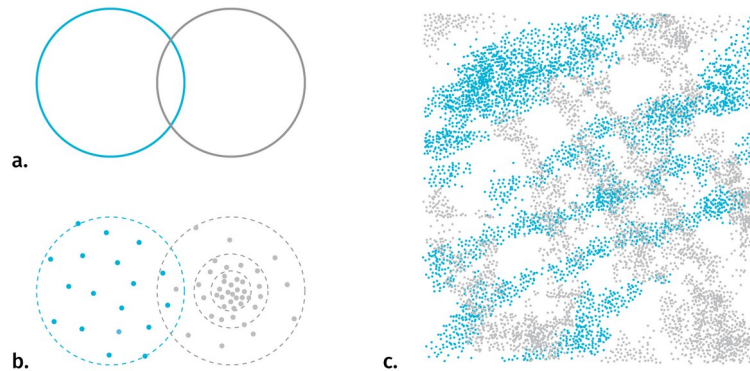


Fig. With polygonal features it is usually straightforward to identify the type of spatial relationship in 2D space (a). When replacing point clouds with polygon representations to apply set logic, the problem of meaningful boundary delineation arises (b). For several complex layers it is hard to say anything revealing about their spatio-temporal relationship (c)

Temporal relations are measures of coincidence. There are thirteen possible relations between two temporal records described in Allen (1984). As we have seen with stations, agents and events, the existence and data collection of any entity can be either continuous or discrete in time, it is therefore useful to distinguish between *time point* and *time interval* when investigating temporal relations (see figures). Linear conceptualization of time can be supported with cyclical and branching time, there can be discrepancies between the temporal parameters of the base map and the thematic overlay, or between the time interval of existence and representations. We'll untangle these complexities in chapter 5.

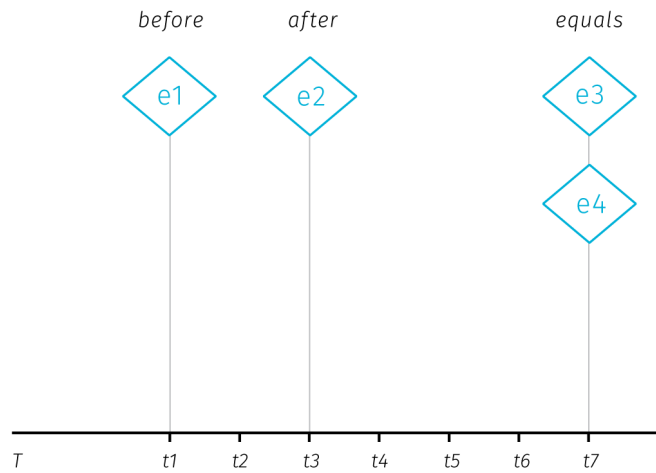


Fig. Temporal relations between time points. Adopted from Aigner, Miksch, Schumann, & Tominski (2011)

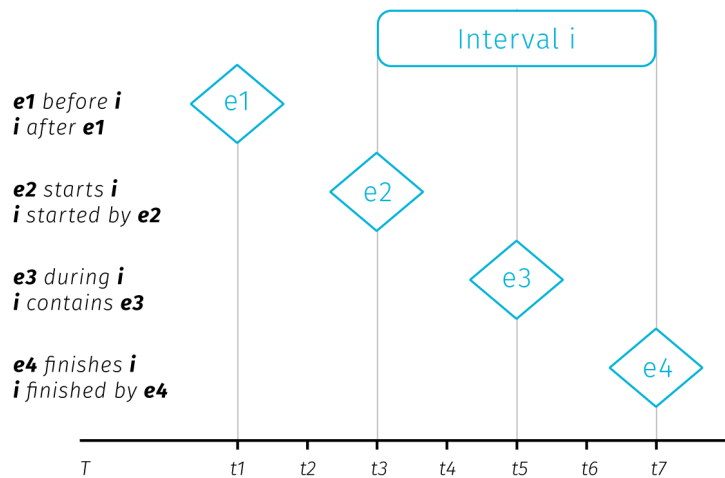


Fig. Temporal relations between time point and time interval. Adopted from Aigner et al. (2011)

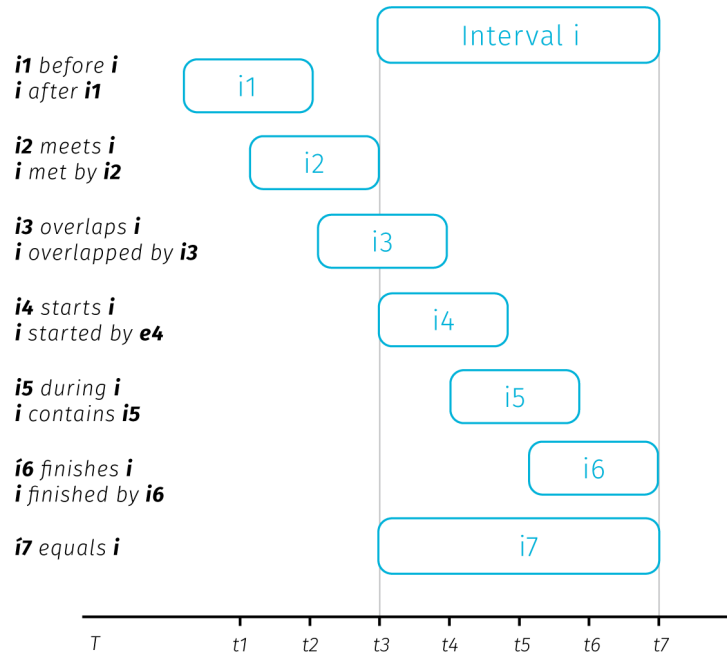


Fig. Temporal relations between two time intervals. Adopted from Aigner et al. (2011)

Relations specific to moving objects – moving objects have a specific set of properties based on their spatio-temporal circumstances. These can be *instantaneous* (actual position and speed), *interval-based* (e.g. travel distance from departure), *episodic* (related to external event) or *total* (related to entire trajectory). (Laube, Dennis, Forer, & Walker (2007), andrienko2008basic).

2.2.2 From data mining to visual analytics

Having described the fundamental spatio-temporal relations in big data sets, we can briefly describe some of the methods to uncover them. Recalling the human-machine continuum at Fig., we will start at the machine side with methods from

the data mining group to eventually move towards the causality interpretation on the human side.

Several data mining concepts are of interest. *Association rule mining* is searching in databases for conditions occurring together frequently. We can describe an association rule as:

$$x \Rightarrow y (s\%, c\%)$$

Where x, y are conditions, together forming an *itemset* and s, c are levels of support and confidence. Support and confidence are basic rule performance measures, support being the measure of how often the itemset occurs in the whole database and confidence being the proportion of x being a member of an itemset $x \Rightarrow y$. For example: *park* \Rightarrow *school* (4%, 55%) means that 55 percent of parks are near schools, for 4% of items in the database (Han, Pei, & Kamber (2011)). The measures of support and confidence allow us to set thresholds for significantly frequent co-occurrence.

Spatio-temporal association rules extend association rules to describe how objects move among a set of regions over time (Verhein & Chawla (2008)). Incorporation of spatiality into association rules takes form of a simple binary conditions telling if the items co-occurred in the same predefined sets of regions or not.

Sequence mining is searching for patterns in time and other sequences. Similarly to association rules, we search for events occurring frequently together by considering three parameters: the *duration* of the whole sequence, the *event window* (time-horizon for considering events as temporally coincident) and the *time interval* between events (Miller (2015)). These parameters allow us to turn the temporal relations between two items into binary parameter telling if the items co-occurred (that is when the time interval between them fits into the event window).

Periodic pattern mining is a type of sequence mining that searches for recurrent patterns in time sequences. Such patterns can be: *full periodic patterns*, *partial periodic patterns* (e.g. just on Mondays), and *cyclic or periodic association rules* that associate events that occur periodically together (Han et al. (2011)).

Considering the breadth of possible spatial and temporal relations described earlier, the conceptualization of spatial and temporal co-occurrence in the association rules may seem rather simplistic. Basically, it is reduced to a yes/no parameter. Moreover, moving from the level of individual database entries towards assessing relations between compound entities such as spatial point clusters seems to be out of the scope of these methods. Of course, the way how spatiality is inscribed into association rules could be made more sophisticated, though with inevitable implications for mining performance. With large datasets, mining even the simple rules forces us to consider time constraints. For such tasks, a simple visual exploration is more efficient and reliable than basic algorithmic solutions.

At this point we can step back from mining algorithms to invite some human interpretation and to consider what conclusions we can actually draw from spatial and temporal co-occurrence of events. The usual assumption is that such co-occurrence can point to some form of causality. Drawing from approaches by Allen, Edwards, & Bédard (1995) and Galton (2012); Bleisch, Duckham, Galton, Laube, & Lyon (2014) distinguish between the trigger that apparently causes the event and the environmental conditions that have to be fulfilled for the effect to occur.

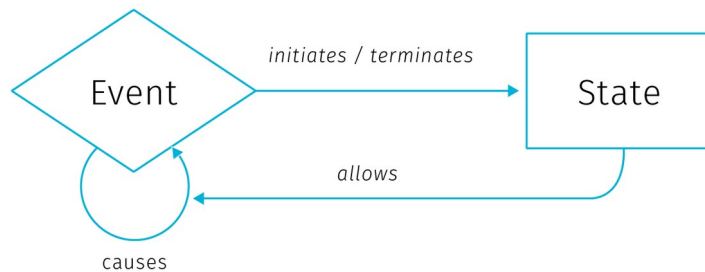


Fig Ontological model of causation, adopted from Galton (2012).
Description in text.

In this model, *state* is an environmental condition and *event* is a change of state. Events are caused only by other events, while states only affect causation by allowing events to cause other events. Events *initiate* and *terminate* states, while states *allow* causation. The *initiate*, *terminate* and *allow* relationships are then dubbed *causal-like* to distinguish them from the event-to-event causation.

In conceptual framework for finding *candidate* causal relationships in movement patterns Bleisch et al. (2014) distinguish between three kinds of granularity at which we can describe phenomena: *spatial*, *temporal*, and *causal*. While the first two are defined by the smallest spatial and temporal units, causal granularity is given by the kinds of events observed. Spatial and temporal granularities can be easily reduced to “see the bigger picture” (by changing the spatial scale, or extending the time range of observation), but causal granularity is more firmly determined by the data collection design.

El-Geresy, Abdelmot, & Jones (2002) note that although the general expectation would be that the effect occurs immediately after the cause, some delay between the effect and the cause can occur, possibly because the cause must attain some intensity threshold to trigger the event or

because the effect and cause are spatially separated and it takes time until the influence of the cause reaches the location where it takes effect. Bleisch et al. (2014) suggest that these apparent delays result from lower causal granularity of observation, i.e. there is some intermediary chain of effect and cause that happens during the delay but it is not recorded by the observation. Whether we accept the effect delays as real or illusionary might be more of an academic question, tracing down the potential causal link between the initial and the final event can yield predictive potential even when the intermediary causal chain remains undiscovered.

Discussing the interpretation of spatio-temporal co-occurrence we have moved on the human-machine continuum towards the human end. At this point, visualisation becomes important as an interface between the user and the data. One of the general models describing how knowledge discovery proceeds via inference and interaction is the sense-making loop (fig).

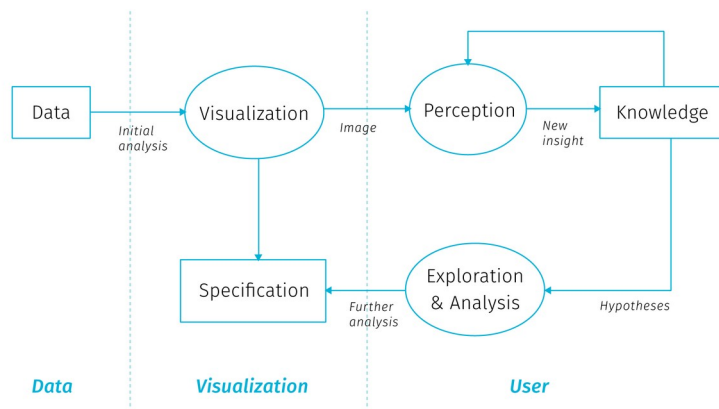


Fig The sense-making loop for Visual Analytics, adopted from Van Wijk (2005). User can interactively manipulate the visualisation to gain understanding of both the data and the representation itself.

Visual analytics extends the concept of visualisation: not only it provides a visual interface to the database, but also makes the data processing pipelines transparent for an analytic discourse. Keim2008visual in their introductory paper say the goal of visual analytics is the creation of tools and techniques to enable people to:

- Synthesize information and derive insight from massive, dynamic, ambiguous, and often conflicting data
- Detect the expected and discover the unexpected
- Provide timely, defensible, and understandable assessments
- Communicate assessment effectively for action

This is truly a long way from the low-level search for co-occurrences, though it is not clear how should these grand goals materialize in practice. Keim2008visual call for broad inter-disciplinary collaboration between related fields (Visualisation, Data Management, Data Analysis, Perception and Cognition, Human-Computer interaction) and identify a range of application and technical challenges.

The brief tour we just went through lets us appreciate the prospect of gaining the best of the both worlds — to support human analytical efforts with algorithmic power doing the heavy lifting around data manipulation. We have seen that inscribing spatiality and temporality to data mining processes can be both cumbersome and simplistic. Furthermore, the co-occurrence we want to search for needs to be defined beforehand, so in many cases data mining is insufficient to provide the required insight. Search algorithms can be performance heavy, which invites some coordination with human observer that is able to easily gain an overview of clusters beyond individual database entities. Visualization and visual analytics provide this exploratory potential, especially for big data in situation where we don't

yet know what questions we want to ask. Visualisation as a sense-making tool gives us a way to find things that we had no theory about and no statistical models to identify and to explore the space of models in more expansive ways (Bollier & Firestone (2010)).

Many possible data transformations may be applicable to a particular problem, but it is not necessarily clear which ones will be of most value in facilitating insight. Also, because visual analytics is qualitative as well as quantitative, there are no assumptions of exact parameters and well-defined boundaries between what is interesting and what is not. A priori criteria of significance may be manipulated based on the judgment of the analyst (Thomas & Cook (2005)). As we will see next, digital cartography has great potential and means to dynamically support cognitive tasks in the manner of visual analytics.

2.3 The role of cartography

Cartography has a long tradition of making data comprehensible to our visual minds. Beautiful and authoritative maps in school atlases explaining for example the formation of air masses or the flows of ocean streams give off an impression of exactitude and definitiveness. But the fact is that these maps are based on data from loads of observations. These data had to be collected, brushed and analyzed for the presence of meaningful patterns, and then visualised in a way that would appeal to human comprehension. The process for creating such maps is nowhere near “real-time” but allows for fine-tuning of all aspects of the map: from carefully shading the outlines of water bodies to making the street connections visually pleasing. This process allows for perfectionism, and the resulting maps remain beloved by collectors long after their ‘utilitarian’ function is gone.

For digital cartography⁵ it took a long time to come any closer to the visual quality of the best works of cartography in print. Arguably, there is still some unfulfilled potential in getting towards graphic excellence in web mapping, though recent improvements in available tools open new possibilities for innovation. Digital maps have the obvious advantage of allowing interaction – user can zoom, pan, change, filter and combine the displayed data. The second big advantage is the possibility to update the displayed data real-time as the data source is updated. Sure, many digital maps are not dynamically updated, simply because the theme does not require it (e.g. medieval monasteries in France or 1991 election results in Yugoslavia). But interactive maps based on dynamically updated data are of special interest as they pose a whole new set of challenges to authors. Ensuring cartographic quality now means designing for yet unseen changes in data with user-induced modifications in mind.

School atlases serve as a presentation of knowledge, are *confirmatory*. Digital cartography allowed for *exploratory* mode of map interaction to emerge, or more precisely, the data exploration step moved from *before* to *after* map publication, and from the cartographer/author to the map user. Visual analytics based on spatial data provides interfaces to manipulate and visualize information, or better to say to pick from the pre-designed visualisation modes. This has implications for both the cartographer and the user.

In the following few sections we will describe what kinds of inference digital cartography aims to support, then we will

⁵ For brevity, we will use the term *interactive maps* as a shorthand referring to maps based on dynamic data, allowing user interaction, consumed almost exclusively through the web, viewed on screens of various sizes. With the term *digital cartography* we will refer to the theory and practice of creating such maps.

outline big data related research challenges for cartography, as well as issues of collaboration and user engagement.

2.3.1 Maps for answering questions, maps for asking them

Interactive map as a data manipulation interface is useful for those who know what questions they want to ask, but also for those who want to find out what they might be asking. So what kind of inference should an interactive map support?

We can start simple, with basic quantitative questions. A big advantage of interactive maps over print is that we can display the exact quantities on demand (e.g. with some pop-up window bound to cursor hover action) and not rely on the viewer's ability to infer quantities from the legend (especially if categorized to some interval scale). The ability to answer simple quantitative queries shouldn't be left in vain, because as Tufte, McKay, Christian, & Matey (1998) warns: "when scientific images become dequantified, the language of analysis may drift toward credulous descriptions of form, pattern and configuration [...] rather than answer to questions *How many? How often? Where? How much? At what rate?*".

We can say that these questions are at the basic level of map reading. bertin1983semiology distinguishes three reading levels for the thematic map, and at each level, different sorts of questions that can be asked:

- *elementary level* – questions introduced by a single element of the visualisation (What is the level of unemployment in this district?)
- *intermediate level* – questions introduced by a group of elements or categories in the visualisation (What are the five most populous districts in the region?)

- *overall or global level* – questions introduced by the whole visualisation (What are the spatio-temporal trends of traffic in this city?)

It is obvious that even a simple map has a potential to introduce countless possible combinations of questions at various levels. As we will see in the next chapter, showing the basic quantities gets complicated in the context of big data, when the number of records to be displayed precludes displaying them individually. Another challenge comes with multiparametric visualisation, especially if we want to support both elementary and global levels of reading for individual parameters.

Besides the importance of supporting elementary-level questions, in thematic cartography we are often interested mainly in the global level of reading as it is hardly achievable with non-cartographic means. Often times, just to *see* the overall level is a revelation — an overreaching macroscope perspective unique to maps. But what else we can do with the overall patterns?

Are there any examples of cartographic visualisation successfully supporting the analytical reasoning? Maybe the most frequent answer for this question would be the celebrated map of the cholera outbreak in London 1855 by John Snow that helped to identify the source of the epidemics in a polluted water pump. This feat is lauded for launching spatial epidemiology and for bringing the thematic cartography to the fore (Clarke & Pickles (2015)). But what exactly made the Snow's method worth following? Tufte et al. (1998) notes four features:

- Placing data in appropriate context for assessing cause and effect
- Making quantitative comparisons
- Considering alternative explanations and contrary cases

- Assessment of possible errors in the numbers reported in graphics

These characteristics describe Snow's thought process which both resulted in and was guided by the map in the making. Indeed, creating effective visualizations is itself a process of exploration and discovery. Working on an interactive map is an iterative process that often yields new questions about the data that were not asked during the early analysis, which enhances the application for the user's benefit.

Modelling what kinds of tasks can be supported by the data is one of the first steps towards a successful visualisation. As Fisher & Meyer (2017) note, high-level questions need to be refined into specific, data-driven tasks. To do this, we can break down the question into four specific components: objects, measures, groupings, and action. Ability to discern those components is a good indicator of whether the task is specific enough and can be computed from data:

- *Objects*: when a task is specific enough, each object will be something that is represented in data.
- *Measures*: In a sufficiently specific task, the measure is either an existing attribute in the dataset or one that can be directly computed from the data.
- *Groupings (or partitions)*: Attributes or characteristics of the data that separate the data items into groups. In a specific task, partitions are attributes of the objects or can be calculated directly from those attributes.
- *Actions*: Specific operation being done with the data such as compare, identify, characterize, etc. Actions guide the process of choosing appropriate visualizations.

There has to be a traceable path from the high-level abstract questions to a set of concrete, actionable tasks in the map

based application, otherwise some additional data may be needed for the questions at hand.

Maps allow for basic quantitative questions on the elementary level, pattern descriptions on the global level. There is much we can do to quantify the pattern descriptions using GIS tools and geostatistics and we can observe spatial correlations between datasets. Spatial patterns in the real world are rarely independent from the geographic context, usually there some observable non-random tendency or some visible or quantifiable relationship with other data layers.

However, the search for patterns and correlations is not the full picture of the feasible use cases. Searching for outliers is interesting for ventures looking for a unobvious opportunities. Similarly, finding areas where the mapped phenomenon is absent can point to development potential. Another use case is searching for deliberate randomness when illicit actors attempt to operate in a fashion that is not predictable from large datasets (Bollier & Firestone (2010)). And then there is a modeling faculty of digital maps that enables what-if questions and comparison of various scenarios. As we have seen, the types of analysis that maps support is broad and each project can yield its own specific kinds of observations, to which we can adjust the custom-made map applications.

2.3.2 What next? Research challenges

Researchers in cartography and geovisualisation see big data as an opportunity and also as a certain call to action. The research agenda for geospatial big data and cartography laid down in Robinson et al. (2017) shows the general interest of moving the field toward fulfilling its potential to make maps that “pique interest, are tacitly understandable and are relevant to our society”. It is certainly reassuring that the

community is aware that new sources of data “stretch the limits of what and how we map”. Building on this, Robinson et al. (2017) list several large-scale and long-term research challenges to face cartography in relation to big data as well as some short-term research opportunities for more concentrated investigation (see appendix A for the overview). Even though some of the points seem vague or repetitive, and the influence of the distinct ICA⁶ commissions is clearly visible, the agenda states some truly exciting challenges to tackle. In relation with the scope to this thesis we can highlight the following challenges for cartography:

- *Develop visual analytical reasoning systems that can help users add meaning to and organize what they discover from geospatial big data* – we need to move beyond naive exploration and focus attention on tools that help people reason about what they are seeing. Users need to be able to save, annotate and compare their findings as they work on complex problems.
- *Develop methods that embody the volume of geospatial big data* – we need cartography that can intelligently process and display big data at a size and a format that users can realistically handle. This will require solutions that support coupled analysis and visualisation as big data often need to be analysed before they are visualised (the order is reversed in exploratory visualisation).
- *Create maps and map-oriented interfaces that prompt attention to important changes in dynamic geospatial big data sources* – We will need to work with global changes, local changes and combinations across scales. In addition, if we display every possible change at once, then the graphical displays become cluttered. Creating summaries of change may be the solution, but we do

6 International Cartographic Association

not yet know how to select important patterns and generalize to something that a user can understand.

- *Leverage what we know about map animation and interactive cartography to construct visual solutions for dynamic sources of geospatial big data* – Conventional solutions for interactive mapping, animated mapping or geovisual analytics can be used for representing big data. However, because of the high velocity characteristic of big data, it is necessary to develop solutions that can automate map design decisions to support interactive design solutions that respond (or potentially precede based on modelled outcomes) as the data changes.

[TODO later link to sections that resonate with the above goals]

As Thomas & Cook (2005) describes, “an emerging discipline progresses through four stages. It starts as a craft and is practiced by skilled artisans using heuristic methods. Later, researchers formulate scientific principles and theories to gain insights about the processes. Eventually, engineers refine these principles and insights to determine production rules. Finally, the technology becomes widely available. The challenge is to move from craft to science to engineering to systems that can be widely deployed”. Cartography, being a university study field had arguably crossed the four stages in the past, though with constant advances in tools for data processing and building interactive applications, the field could benefit from regularly revisiting the craft stages to see how the new tools alter our concepts of mapmaking.

This thesis does not have the ambition to imagine all the paths cartography could take in the future. However, in addition to mentioned agendas, we would like to highlight three overreaching questions that we feel are not widely discussed within cartography. Much of the work described

in the remaining chapters of this thesis is rooted in pondering on the following questions about the practice of digital mapmaking:

(a) Is cartography fully exploiting the digital medium?

Before hopping on the wagon of augmented reality and immersive experiences (that make roughly a tenth of the population sick) cartographers could consider if they made the most of the previous medium shift. Even in the plain world of regular screens and everyday web traffic there is still a lot to be achieved for cartography to be truly useful for everyone.

Web is inherently a map-friendly platform where map products will be increasingly commonplace. Yet from the cartographic perspective, the great portion of thematic maps on the web seems rather underwhelming. Default-style markers for points of interest and numbered marker clusters to “solve” high point densities are just the tiny portion of what could be done. Cartographers should be the first to go beyond the pre-set graphic means.

Apart from the limitations posed by opinionated mapping frameworks there are also certain mindset limitations that come from transferring a visual artifact from one medium to another. Rules and practices that were to a large degree dictated by the old medium of transmission (print) get involuntary transposed to the new medium that may not require them at all. This was apparent for example in the grid-like organization of the web news pages transferred initially from the design of printed newspapers. Are there such taken-for-granted givens that linger meaninglessly in digital cartography?

There are of course many limitations that are not imaginary, like data interoperability issues and vendor lock-ins. The skill sets needed for data analysis, desktop GIS

operation and web development seem painfully detached. But all the problems apart, a good mental exercise for cartographers would be to imagine map creation and interaction detached from any medium – what would we design if anything was possible?

In cartographic research, we often test the cognitive efficiency of the visualisation methods that already exist. This is all good, but we should not assume that the cartography's quest to *extend* the arsenal of visualisation and interaction methods is completed. As we will see further, interaction and increasing data load pose new challenges to cartographic visualisation, with opportunities for creative inclusion and combination of new methods.

(b) What inspiration can digital cartography take from the heritage of pre-digital mapping?

Same as we asked about the preconceptions of the old medium, we can reverse the question and ask if there are any good tricks from the rich history of cartography that did not make it to digital mapping toolbox. What was lost in transition to digital? Even though paper maps and atlases age in the sense of content, cartographic methods used in them often remain inspirational and valid — the old map products many not be outdated for cartographers.

Some of the classical cartographic techniques may be demanding to implement in the variable scale environment, same other may provide solutions to the visualisation issues like high density displays. Again, we arrive at the problem of opinionated web mapping libraries that are not easy to customize or extend. Cartographers usually aren't software developers, and software developers are usually unaware of old map stocks, but there are already examples of positive trends in collaboration towards richer visualisation in digital maps.

(c) Should cartography focus more on the interaction design?

Creating digital maps is not only about assigning appropriate visualisation type to the data at hand. It also becomes increasingly about designing user interactions with map elements. The ways how the map application enables user actions, the way how map controls and map elements react to user-induced changes, the way how the whole map composition adapts to screen space constraints, this all weaves a complex net of interdependent design decisions that will become an inherent part of digital cartography. What is more, the challenges of high data density affect both map intra-composition as well as extra-composition⁷.

As Robinson et al. (2017) note, too often in the visual analytic process, researchers tend to focus on visual representations of the data but interaction design is not given equal priority. We need to develop a “science of interaction” rooted in a deep understanding of the different forms of interaction and their respective benefits.

In the IT industry there is a discipline of UX (user experience) that could provide some inspiration for example in accessibility evaluation, though most of its methods doesn't fit very well to the specifics of interactive maps. At the same time, large web map providers probably collect user interaction data that could power cartographic research if they were accessible.

But we shouldn't limit our sight to software interfaces to get inspiration. There are tons of well designed devices in the physical world that could serve as an example of clever interface design. There is a potential on expanding the repertoire of interaction techniques for digital maps. As mapped themes vary greatly the interactions could be tailor-fit as is often is with visualisations. As we have seen many

⁷ TODO explain these terms?

times in history of innovation, progress is often hampered by mental roadblocks we don't even realize we have.

2.3.2 How next? Challenges in collaborative practice

Having described the ontological models of causation as well as visions for the future of cartographic research, we can now take an aside to dwell a bit on the nitty-gritty realities of mapmaking in practice. Practical aspects of the profession are often overlooked in literature, as well as the fact that cartographer often needs to operate within a greater team. The smoothness of collaboration within the team is then a determining factor of its productivity.

There is a (somewhat mythical) notion of “full-stack” visualization designer-developer capable of conducting the broad range of tasks needed for a visualisation project (Gray, Chambers, & Bounegru (2012)). Though some such individuals do exist (possibly working on smaller applications for customers or on PhD projects), it is clear that cartographer can take only so much of additional roles (data analyst, UX designer, front-end developer, database administrator...) before getting on thin ice. Real-life visualisation projects often include a range of team members or even teams with disjoint skill sets. The question then arises on how to modularize the work. One possible model of decomposition is the information visualisation reference model (**Fig**).

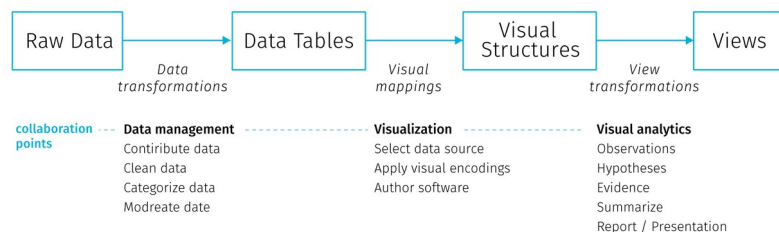


Fig Information visualization reference model. Adopted from Heer & Agrawala (2008)

In this model the collaboration points lie at the transitions between the stages and involve decisions on data management, visualisation and analytical capabilities (Heer & Agrawala (2008)). Physical and temporal separation of teams and institutional and disciplinary divides lead to early-stage partitioning of tasks both in the *design* (data profilation, ideation, mockup creation and prototyping) and *development* (implementation, testing, deployment and maintenance) phase (Walny et al. (2019)). Such divisions are not unique to data visualisation projects, it could match any web development project.

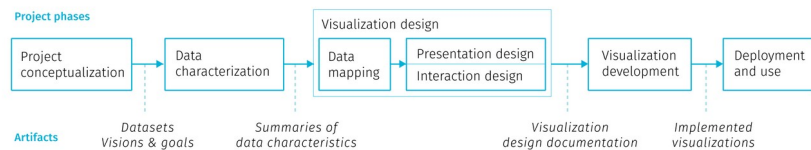


Fig Stages of data visualization development process. Adopted from Walny et al. (2019)

Walny et al. (2019) formalize stages of data visualisation process based on experience with several assignments (Fig.). In the iterative process the division of labor gives rise to *handoff* events, when one team passes work products and requirements to the next team. Particularly the handoff between the design and development team is where issues can arise to affect the end result. Speaking from the position of design team Walny et al. (2019) articulate several key challenges that affect the success of the handoff and in turn the smoothness of the whole project:

- *Adapting to data changes* – changes in input data can have cascading effects throughout the stages of the process. Some breakages are inevitable (e.g. API changes) and fixing them is a part of project

maintenance. It is advisable to have data transformations automated to the largest extent possible, as it is highly likely there will be a need to reiterate them. In this sense, the scripts and the processing tool chain developed during the project can be actually more valuable to creators than the project outputs.

- *Anticipating edge cases* – though this is incredibly hard for real-time data inputs, best effort should be made to foresee at least the main application states resulting from the user interactions, such as filtering, changes of scale, etc.
- *Understanding technical challenges* – knowledge of technical constraints helps to produce feasible design ideas. Development team's concerns differ from the design team's, they include cross-browser compatibility or future code maintainability. In some areas the goals can overlap, for example in accessibility considerations or performance optimization.
- *Articulating data-dependent interactions* – prototyping interactions such as linking and brushing using conventional graphic tools is challenging, not to mention animations or transitions between views. There are wireframing tools that try to address this, though misunderstandings still occur.
- *Communicating data mappings* – this is a concern when delivering static mockups for the development team. The mapping between data and the interface controls may not be obvious, especially when the complexity of data does not allow to exemplify all possible application states. Annotations within mockups try to ease this.
- *Preserving data mapping integrity across iterations* – tracking implementation adherence to the design, finding errors, as well as checking if change requests from previous iterations got implemented is solely a matter of visual inspection and therefore prone to error.

This can be fixed by automated testing, though it is not feasible for all types of projects, and even if implemented, the test coverage can rarely reach 100%.

Project examples from Walny et al. (2019) show that the above challenges were formulated based on project experience with relatively static data inputs. This underlines why interactive cartographic visualisation of real-time data is hard: much of the advice is hard to impossible to follow when the real-time data inflow is volatile. [TODO link on chapter on data mocking]

2.3.3 Who cares? Building user engagement

The ability to interact with the map-based application can surely be empowering for the user, triggering the sensation of exploring the unknown. On the other hand, things can go wrong as it is very hard to create an immersive experience from a complex dataset that would be immediately understandable to the newcomer. Exploratory map applications intended for the general public can leave users overwhelmed with the amount of possible interaction points. Left to their own devices and without any stated framework for interpretation, users need to create their own narration about what is displayed. Visual interfaces are prone to be terrifyingly cluttered, untroubled with dangers of fostering misinterpretation. Lack of guidance on where to start results in poor engagement with the application that is quickly abandoned. With specialized applications for professional audience, this can be mitigated by training, because users are basically forced to work with the application as part of their job. Similar problems occur in business analytic dashboards proliferating in enterprises, which fail to make sense to users, or worse, fake insight with vaguely understood and hardly interpretable metrics. All these caveats pose a challenge to application designers.

Building engagement with an application is mainly about sparking curiosity in users. Previously in this chapters we discussed what kinds of questions can interactive maps answer, assuming that there is someone who wants to ask them. But this assumption shouldn't be taken for granted. Adding hints and motivations to map interfaces should get more interest from cartographers, not only because it is a way to make our work resonate in general public but also it provides some options to work around the cognitive difficulties connected with visualizing large amounts of data.

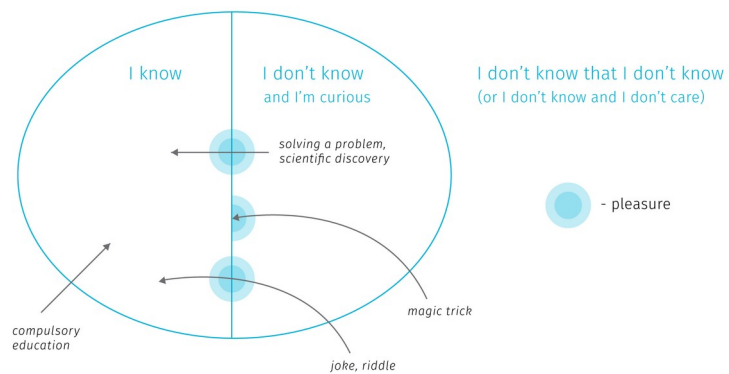


Fig. Various ways of expanding our mental model or the world can trigger pleasure. Adopted from Stanová (2016)

So how to trigger curiosity? According to Stanová (2016) (Fig), to set off the thrill, one needs to first get from the “I don’t know that I don’t know” zone to the “I don’t know and I’m curious” zone. Here we realize that there is something missing in our mental model of the world and we want to discover it. Crossing the border from “I don’t know” to “I know” creates pleasure and motivates further exploration (which is crossing back and forth between “I don’t know” and “I know”). Note that the “I know” zone doesn’t need to correspond to reality — exploring fictional worlds, gossips or conspiracy theories triggers pleasure just as well.

From the cartographic point of view, when we create maps for specialized audience, we target people who already are in the “don’t know + curious” zone. But when working for general audience, it is important to think about how to move users to the curious zone to encourage interacting with an application. Sometimes the compulsory education makes the mistake of revealing how something works before making pupils interested in the topic, resulting in low levels of engagement and remembering. The same problem can occur when the user lands straight at a complex web interface with intricate cartographic visualisation.

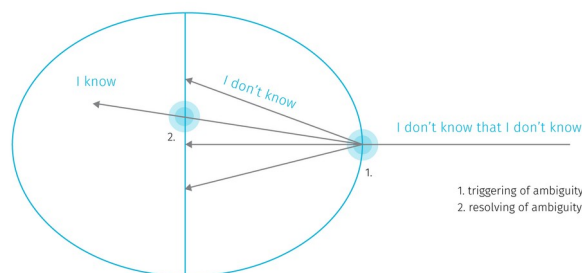


Fig. Pleasure points usually arise when ambiguity is triggered and when it is resolved. Adopted from Stanová (2016)

The thrills can arise also when crossing the I don’t know barrier, that is when the ambiguity is triggered (Fig). The pleasure here comes from the expectation of the potentially upcoming reward — resolving the ambiguity. However, note that the ambiguity doesn’t need to be resolved to set off thrills — for example, when watching a magic trick we don’t get to know how the magician does it but we can entertain ourselves in speculations. Similarly in exploratory cartographic interfaces the stimulation comes from exploring various aspects of the displayed data, fostering our curiosity about the nature of the phenomenon.

The related concept of *flow* describes the relationship between the skill of the user and the difficulty of the task at hand (Csikszentmihalyi (1997)). Flow is the mental state resulting from the right balance between the task difficulty and the user's skills. The flow channel is rather narrow — if the difficulty is too high, anxiety arises, if it's too low, user is bored. The task of the interface designer is not to dumb down the displayed content, but also not to create additional roadblocks with incomprehensible interaction modes.

Shaping the initial experience with an application is sometimes referred to as “user onboarding” (Baur (2017)). The onboarding stage is far too often omitted in information visualisations on the web, though there is no wide consensus on how it should work, the basic aim should be at a minimum a short textual introduction to the presented data and the knowledge around it. Complex interfaces tend to offer an initial tour of controls that usually explains what individual UI elements do, less often it ventures into teaching people how to read and understand the presented visualizations. There are approaches like “scrollytelling” (Amabili (2019)) or explorable explanations (Victor (2017)) that embed interactive visualisations into larger body of text that gradually explains it.

The task is further complicated by the obvious fact that different people have different interests and understanding of the world. In web development, creators try to specify user personas to represent target audience and user stories to list supported user actions. To start modelling from considering users rather than data is certainly an advisable approach. However user stories are often too removed from the real life circumstances to perceive and prevent abusive uses of the system (so called “weaponisation” of design Diehm (2018)).

Tailoring the map interaction to intended users is definitely something digital cartography should aim for. While some customizations are technologically possible (e.g. the dark mode in transportation maps switching on for tunnels or night travel), the topic of accessibility of interactive maps is still fairly unexplored. While there are methods to ensure map is usable for people with color blindness, there is a wide range of vision impairments cartography is largely unprepared to adjust for.

2.4 Objections and Responses

Not surprisingly, data visualisation is seen as a great tool for achieving a desirable goal in most of the literature, this thesis included. But there are also voices raising objections against some naive expectations from data visualisation that may even get harmful with incorporation of big data. Such debates may shed some light on the possible future evolution of digital cartography, so let us briefly outline some of the main objections with possible responses to them.

Hiding system complexity

The first objection is related to modelling and visualizing complex systems. By reducing the complex system into comprehensible chunks, data visualisation may encourage confident predictions and estimations that may lead to decisions and interventions with harmful effects. The issue is that the used datasets and models may not include all aspects of reality, but absence of evidence is not evidence of absence. Incomplete models may rationalize decisions that despite claiming to be data-driven, actually stand on thin legs. This can lead to unpleasant surprises (for example, loosening of pandemic-related restrictions in Czech Republic before Christmas 2020 was based on models that did not

account for the presence of a COVID-19 variant with increased transmissibility, leading to disastrous effects soon afterwards). Visualisation based on a bad model than acts basically as an accomplice, no matter how well crafted it is.

Apart from missing relevant information in models, there are some aspects of complex systems that render prediction efforts highly problematic. Nonlinear relationships between system parts produce disproportional responses to change in input parameters. The boundaries between system components may be hazy or even imaginary. When the relationships between system components are vaguely understood (which is often the case in complex systems), any naive intervention can trigger a chain of cascading second order effects that can accelerate harm (Meadows (2008), Taleb (2012)). Moreover, extracting trends from historical data offers no preparation for the so called Black Swan events — large-scale unpredictable and irregular events of massive consequence (Taleb (2007)).

What can be done in response to these concerns? Models and visualisations could be more explicit about their own limits — like the properties of data used, statistical assumptions, margins of error, sources of uncertainty and possible other explanations. Incorporation of uncertainty into visualisations should be done in a way that cannot be easily ignored by the users (Kale, Kay, & Hullman (2020), Correll, Moritz, & Heer (2018)). Visualisations shouldn't simplify the depiction of data if it leads to hiding important aspects of the system. Overall, visualising interdependencies and feedback loops within a complex system poses an interesting challenge for visualisation community and for cartographers as well.

When it comes to reasoning about complex systems, visualisations can support a non-predictive approach that aims more at risk evaluation and moderating exposure to

the possible harm coming from unprecedented events. In relation to man made systems in general, we could strive to make them more akin to natural systems that are not only robust to error, but can also adapt and benefit from certain doses of volatility (so called antifragile systems Taleb (2012)).

Misinterpretation

The second objection is related to the first one, but rather than looking at limits of models and visualisations, it is concerned with user's ability to interpret them. It is long known that our intuitive thinking is influenced by biases in many tasks, including assigning probabilities to events, forecasting the future, assessing hypotheses and estimating frequencies (Kahneman (2011)). There is a long list of cognitive biases that correct the assumption of people being fully rational actors.

For example, biases about the reliability of different sources may lead us to discount information from sources that we don't associate with (Thomas & Cook (2005)). When we form a preliminary judgment too early in the analytical process, we may hold firm to it long after the evidence invalidates it. Sometimes we settle for a "good enough" answer, stopping our analytical process before identifying critical information that would lead us to a different conclusion (Heuer (1999)). We are also challenged to think statistically compared to our abilities to think associatively, metaphorically or causally. Furthermore, there is our overconfidence in what we think we know, and inability to acknowledge the full extent of what we don't know (Kahneman (2011)). We also underestimate the role of chance in events, we tend to assume causality between events that just happen to occur at the same time (Taleb (2012)), and so on, and so on.

To make the matters worse, experts who create visualisations are to susceptible to various biases. The design choices can drive results every bit as much as traditional

“data-cleaning” choices. Hence visualization techniques contain embedded judgments (Bollier & Firestone (2010)). Then there is the “curse of knowledge”: the difficulty in imagining what it is like for someone else not to know something you know (Pinker (2015)). On the side of the application designer, it can lead to expecting the user to have same levels of skills and knowledge, but also the same values and views of the world.

What can be done in response to these concerns? First of all, we might benefit from a more realistic view of the impact of data visualisation. Insight comes from knowledge and experience and no data-driven tool can compensate for the lack of the two. Visualisation is a supporting tool, a mediator that can stimulate and amplify the thought process, but cannot act as a shortcut if no thinking is being done. Expecting to become an expert by looking at a picture is a false promise.

That being said, acknowledging the cognitive biases in the visualisation design process is definitely a right way to go. Visualisation designers have often little information on what judgements are triggered by their work. There is a growing body of research on user interactions with complex visualisation that could help us. Applications could incorporate tools to collect feedback from users, even evaluate usage data to find issues. Overall, judgement biases are systematic errors, therefore to some extent they are predictable, although maybe not preventable. The task is then to explore how to adjust the visual language to warn users that they might be biased.

Non-human decision makers

The third question is directed more at the future relevance of data visualisation in the face of artificial intelligence (AI). If more and more decisions will be carried out by algorithms the need for visualisations may diminish in many areas

where it is deemed crucial nowadays, merely because computers, unlike human analysts, don't need to visualize things to gain insight and understand the problem.

Having described the range of cognitive biases one might welcome computational assistance, and in many areas we already rely on it, navigation being a prominent example. On the other hand, we can point out the current deficiencies of machine learning algorithms. Contrary to popular beliefs, the technology is not ready to step in for humans for the majority of tasks. Current machine learning is about extracting rules from vast training data sets, which is susceptible to various kinds of issues: sensitivity to gaps and errors in data, confinement to the specifics of training data, or the tendency to take unwanted shortcuts (Shane (2019)). Furthermore, it is not certain that some breakthrough in artificial intelligence will come in the foreseeable future, either because its already large appetite for computing power will become economically, technically, and environmentally unsustainable (Thompson, Greenewald, Lee, & Manso (2020)), or because the demand for artificial collaborators will perish — we will simply want to design AI as tools not as collaborators (Dennett (2017)).

In a realistic view, the danger of AI is not in usurping us, but rather in us putting too much confidence into uncomprehending tools. But even if we will get to non-human decision makers, it wouldn't mean a demise of visualisation, contrary, it could open new opportunities for using visual artefacts. First, as a communication interface between humans and machines. Second, as a way to inspect and verify the workings of decision-making algorithms. It is not possible to check upon a black box, nor to negotiate with it, so gaining insight into how and why automated systems arrived at a particular decision may become a new frontier for the data visualisation community.

Aigner, W., Miksch, S., Schumann, H., & Tominski, C. (2011). *Visualization of time-oriented data*. Springer Science & Business Media.

Allen, E., Edwards, G., & Bédard, Y. (1995). Qualitative causal modeling in temporal gis. In *International conference on spatial information theory* (pp. 397–412). Springer.

Allen, J. F. (1984). Towards a general theory of action and time. *Artificial intelligence*, 23(2), 123–154. Elsevier.

Amabili, L. (2019). From storytelling to scrollytelling: A short introduction and beyond. Available online at <https://medium.com/nightingale/from-storytelling-to-scrollytelling-a-short-introduction-and-beyond-fbda32066964> (last accessed April 29, 2021).

Baur, D. (2017). The death of interactive infographics? Available online at <https://medium.com/@dominikus/the-end-of-interactive-visualizations-52c585dcfcfb> (last accessed April 29, 2021).

Bleisch, S., Duckham, M., Galton, A., Laube, P., & Lyon, J. (2014). Mining candidate causal relationships in movement patterns. *International Journal of Geographical Information Science*, 28(2), 363–382. Taylor & Francis.

Bollier, D., & Firestone, C. M. (2010). *The promise and peril of big data*. Aspen Institute, Communications; Society Program Washington, DC.

Clarke, V., & Pickles, R. (2015). *Map: Exploring the world*. Phaidon Press Limited.

Correll, M., Moritz, D., & Heer, J. (2018). Value-suppressing uncertainty palettes. In *Proceedings of the 2018 chi conference on human factors in computing systems* (pp. 1–11).

Csikszentmihalyi, M. (1997). Flow and the psychology of discovery and invention. *HarperPerennial*, New York, 39.

Dennett, D. C. (2017). *From bacteria to bach and back: The evolution of minds*. WW Norton & Company.

Diehm, C. (2018). On weaponised design. Available online at <https://ourdataourselves.tacticaltech.org/posts/30-on-weaponised-design/> (last accessed September 16, 2018).

Egenhofer, M. J., & Franzosa, R. D. (1991). Point-set topological spatial relations. *International Journal of Geographical Information System*, 5(2), 161–174. Taylor & Francis.

El-Geresy, B. A., Abdelmot, A. I., & Jones, C. B. (2002). Spatio-temporal geographic information systems: A causal perspective. In *East european conference on advances in databases and information systems* (pp. 191–203). Springer.

Fisher, D., & Meyer, M. (2017). *Making data visual: A practical guide to using visualization for insight*. " O'Reilly Media, Inc."

Galton, A. (2012). States, processes and events, and the ontology of causal relations. IOS Press.

Gray, J., Chambers, L., & Bounegru, L. (2012). *The data journalism handbook: How journalists can use data to improve the news*. " O'Reilly Media, Inc."

Guo, D., Chen, J., MacEachren, A. M., & Liao, K. (2006). A visualization system for space-time and multivariate patterns (vis-stamp). *IEEE transactions on visualization and computer graphics*, 12(6), 1461–1474. IEEE.

Han, J., Pei, J., & Kamber, M. (2011). *Data mining: Concepts and techniques*. Elsevier.

Heer, J., & Agrawala, M. (2008). Design considerations for collaborative visual analytics. *Information visualization*, 7(1), 49–62. SAGE Publications Sage UK: London, England.

Heuer, R. J. (1999). *Psychology of intelligence analysis*. Center for the Study of Intelligence.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kale, A., Kay, M., & Hullman, J. (2020). Visual reasoning strategies for effect size judgments and decisions. *IEEE Transactions on Visualization and Computer Graphics*. IEEE.

Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., & Melançon, G. (2008). Visual analytics: Definition, process, and challenges. In *Information visualization* (pp. 154–175). Springer.

Laube, P., Dennis, T., Forer, P., & Walker, M. (2007). Movement beyond the snapshot–dynamic analysis of geospatial lifelines. *Computers, Environment and Urban Systems*, 31(5), 481–501. Elsevier.

Meadows, D. H. (2008). *Thinking in systems: A primer*. Chelsea green publishing.

Miller, H. J. (2015). Spatio-temporal knowledge discovery. *Geocomputation: A Practical Primer*. SAGE Publications Ltd, Thousand Oaks, CA, 97–109.

Pinker, S. (2015). *The sense of style: The thinking person's guide to writing in the 21st century*. Penguin Books.

Robinson, A. C., Demšar, U., Moore, A. B., Buckley, A., Jiang, B., Field, K., Kraak, M.-J., et al. (2017). Geospatial big data and cartography: Research challenges and opportunities for making maps that matter. *International Journal of Cartography*, 1–29. Taylor & Francis.

Shane, J. (2019). *You look like a thing and i love you: How artificial intelligence works and why it's making the world a weirder place*. Voracious.

Stanová, M. (2016). *Algorithms in art*. Academy of Fine Arts in Prague & CEE PhotoFund.

Taleb, N. N. (2007). *The black swan: The impact of the highly improbable* (Vol. 2). Random house.

Taleb, N. N. (2012). *Antifragile: Things that gain from disorder* (Vol. 3). Random House Incorporated.

Thomas, J., & Cook, K. A. (2005). Illuminating the path: The r&D agenda for visual analytics national visualization and analytics center. *National Visualization and Analytics Center-US Department of Homeland Security*.

Thompson, N. C., Greenewald, K., Lee, K., & Manso, G. F. (2020). The computational limits of deep learning. *arXiv preprint arXiv:2007.05558*.

Tufte, E. R., McKay, S. R., Christian, W., & Matey, J. R. (1998). Visual explanations: Images and quantities, evidence and narrative. *Computers in Physics*, 12(2), 146–148. AIP Publishing.

Van Wijk, J. J. (2005). The value of visualization. In *VIS 05. IEEE visualization, 2005*. (pp. 79–86). IEEE.

Verhein, F., & Chawla, S. (2008). Mining spatio-temporal patterns in object mobility databases. *Data mining and knowledge discovery*, 16(1), 5–38. Springer.

Victor, B. (2017). Explorable explanations. Available online at <http://worrydream.com/ExplorableExplanations/> (last accessed April 29, 2021).

Walny, J., Frisson, C., West, M., Kosminsky, D., Knudsen, S., Carpendale, S., & Willett, W. (2019). Data changes everything: Challenges and opportunities in data visualization design handoff. *IEEE transactions on visualization and computer graphics*, 26(1), 12–22. IEEE.

Worboys, M. F., & Duckham, M. (2004). *GIS: A computing perspective*. CRC press.