

Emergency Information Visualisation



Hoang Long Nguyen and Rajendra Akerkar

1 From Big Emergency Data to Visualisation

We can leverage big emergency data as the standards and guidelines for emergency management, some of which are smart response systems, early event anticipation, and coordination between end-users (e.g., the connection between authorities, first responders, and local citizens). Emergency data is available in various sources (e.g., social media, mass media, sensor, and linked open data) or local knowledge. In terms of the 6 Vs of big emergency data (i.e., value, volume, velocity, variety, veracity, and variability), they are challenging actually and consume our research efforts in analysing and capturing essential knowledge, especially in urgent circumstances. With manual methods, it may take years for a person to read, process, and understand massive amounts of big emergency data [23]. We, therefore, need to come up with a digital solution, which can deal with the complexity of emergency data, to get meaningful results in a fast response time.

Data visualisation is an excellent approach to solve the problems mentioned above by representing either tangible or intangible emergency data in systematic forms without losing any crucial information [8]. With an enormous amount of complex emergency data, our human brain can understand and process visual information (e.g., diagrams, graphs, or charts) 60,000 times faster than raw documents, spreadsheets, notes, numerical tables, or reports [27]. To identify and deal with an image, we need to spend only 13 ms.¹ Data visualisation is a powerful tool to discover latent correlations and unexpected relationships, which is impossible by

¹<http://news.mit.edu/2014/in-the-blink-of-an-eye-0116>.

H. L. Nguyen (✉) · R. Akerkar
Big Data Research Group, Western Norway Research Institute, Sogndal, Norway
e-mail: hln@vestforsk.no; rak@vestforsk.no

manually exploring or traditional descriptive statistics [14]. It can bring us surprising patterns that we may never imagine.

One of the key discoveries in science that disease was passed through germs, rather than through pollution resulted from perception derived from a visualisation of the location of London cholera outbreaks near the Broad Street pump. The visualisation of cholera outbreaks changed how we saw a disease. Present-day Information visualisation can be used to create similar ideas: understanding the spread of a COVID-19 pandemic, spotting terrorist movement, or evaluating the emergency impact on a town. But, there are some notable challenges: several things may be utilised and visualised, often derived or aggregated from vast data sets, or created by algorithms for analytics.

Moreover, we can leverage data visualisation for performing prediction [1] and feature selection [29] in machine learning and artificial intelligence. We can do the data visualisation even before, during, or after data analysis. Visualising data before and during data analysis provides better insights into our data. It supports discovering underlying knowledge quickly and comprehensively; for example, conducting time-series visualisation may help recognise abnormal patterns and spot trends on which we should focus. On the other hand, performing visualisation after analysing data may make it easier, quicker, and more precise to communicate any results and insights to audiences. Even though data visualisation is enormously useful, but inappropriate methods can drive our decisions in negative ways. Hence, we must deeply understand our data, target audiences, and particular scenarios to select visualisation techniques correctly and carefully for achieving the most powerful insights.

Our motivation in this chapter is to provide crucial knowledge to let readers understand how big emergency data should be visualised to overcome challenges, what visualisation techniques are available to support data analytics, and in which situations we can apply. Based on these critical questions, we address in detailed multiple design strategies for an excellent viewing, different methods to display emergency information in various conditions, and several exercises at the final for understanding how to bring proposed theories into practical problems. The knowledge in this chapter can support a wide range of emergency applications that is vital for our community to derive optimal decision-making criteria and strategies, as our world is becoming more and more complex. Besides, we complete this chapter to be used different end-users, from academic researchers to practitioners, from natural to social science areas, and from inhabitants to authorities. The visualisation of emergency information is a must for any community today to increase their resilience [3].

We introduced the problem and emphasised our motivation for researching about emergency information visualisation in this section. In Sect. 2, we present visualisation design objectives for the sake of efficiency in further sections. Next, Sect. 3 summarises different visualising methods for content-based, geospatial, and temporal information along with details of dashboards. Then, we introduce some pertinent research issues in Sect. 4. Last but not least, we provide practical exercises in Sect. 6.

2 Visualisation Design Objectives

The main objective of the visualisation is to explore, develop and provide insight into a data set. However, there are several essential challenges [28] that data visualisation techniques have to face with, especially towards big data. Some of which are mentioned as follows.

- **Understanding our data:** to create an effective visualisation, it is vital to understand our data comprehensively, for example, manifest and latent context, underlying meaning, and hidden patterns. However, following our data has never been easy. In the era of big data, we have to deal with the sheer volume of both structured and unstructured emergency data. A data type is very diverse, as well. It can be content-based, geospatial, and temporal data. Misunderstanding our data can lead to inefficient visualisation or, also, missing or giving wrong information.
- **Dealing with outliers:** outliers are exceptional values that are inconsistent with other observations in a data set on account of measurement variability, sampling problems, wrong number crunching, natural variation or experimental errors. We should remove outliers from our data set because they may create a degree of complexity and ambiguity, which could be the main problems affecting our visualisation, leading us to make the wrong decision. On the opposite, outlier can contain useful information about abnormal properties of entities. We need practical approaches to tackle with and evaluate on that abnormal values to determine whether we should keep or exclude them. However, detecting, analysing, and assessing anomalous instances might be not always easy. Neural networks, support vector machine, logistic regression, and clustering-based algorithms are basic methods to handle outliers.
- **Fast response time:** high accuracy is essential; however, fast response time is a crucial attribute of a good visualisation as well. There are always challenges of querying the overabundant amount of data in a short time to capture and display useful information. In real-time systems, we have to decide a trade-off between discovering the data accurately and visualising it quickly.
- **Displaying meaningful results:** even we can capture all useful information quickly, displaying significant results is the problem of selecting what to display when dealing with enormous amounts of information. We have to define what type of information is displayed, how to show, and in which order. Stuffing everything at the same place can cause audiences confusingly.

Visualisation design objectives can be considered as good starts for brightening the way of people in avoiding these problems as above mentioned and creating useful visualisations. For transforming complicated scientific information into an expression that is meaningful and helpful towards every user group, visualisation techniques should satisfy the following objectives.

- **Setting your target:** to do a visualisation successfully, the first objective is to acknowledge our end goal. We need to identify which our audience prefers by always asking ourselves what our audience is expecting to derive. There may

have some information that is so obvious and seems to be redundant to us; however, the viewer may need to understand the visualisation comprehensively. Determining an explicit target can guarantee that we do everything purposively, reasonably, and understandably.

- **Choosing the right visualisation:** selecting an inappropriate visualisation may demolish all of your hard efforts. There may have several ways to visualise our data. In this circumstance, it is crucial to follow our determined end goal and choose a visualisation methodology that is best suited to the targets of our work and the requirements of our intended audience.
- **Simple is the best:** this is a critical objective to create an effective visualisation. Remembering not to put so much information in your visualisation. This will make audiences challenging to capture their necessary information. If we make the visualisation too complicated and tedious, the audience is going to spend much time understanding the diagram instead of getting hidden knowledge. This is not what we want to have. Instead of cramming all of the information in the same place, you should categorise them into groups and visualise them in different graphs or diagrams. Do not waste our time on futile decorations. We use labels only if they do not cover any essential information on the chart.
- **Be consistent:** a piece of advice, especially for beginners, is to ensure that every design feature (e.g., the use of shapes and symbols, the choice of colour scheme, the order of items, and the selection of position and font of labels) is consistent within your visualisation. A visualisation guaranteeing the consistency can help readers get an overview at a glance.
- **Easy comparison:** one good motivation when using visualisation is the ability to conduct the comparison. Our visualisation will be more helpful if we can compare valuables in the same diagram or even between different diagrams easily and comprehensively. We then can capture the strengths and weaknesses of variables over space or time-periods.
- **Ensuring the clarity:** last but not least, we must verify that every information is visible. The contrast between colours, especially adjacent ones, must be high enough. The placement of all labels on data points and lines must be easy to determine. If there is any text in our visualisation, we should make sure to create high descriptive text and not over-explain. Every duplicated and redundant information should be removed as well.

In the emergency area, visualisation can enable end-users (e.g., authorities, first responders, and inhabitants) to utilise captured information by presenting it in significant ways. This can enhance the value of information and increase decision-making potential. To deal with complications of emergency information visualisation, we provide adequate techniques for a successful data visualisation in the next section.

3 Visualisation Techniques

In this section, we aim at providing different visualisation techniques corresponding to the types of information that we wish to process which are content-based, geospatial, and temporal information.

3.1 *Content-Based Information Visualisation*

What may we do if we were given a massive raw data set? Content-based analysis can be the first step for providing a deep analysing with a variety of methodologies to capture valuable knowledge, to discover hidden patterns, and to identify complex relationships. It is also known as content-based mining and knowledge discovery. In the field of emergency, we can apply content-based analysis for anticipating risks, determining impacts, and comprehending triggered actions. The integration of content-based analysis and visualisation can convert from structured and unstructured data (e.g., corpora, spreadsheets, and social media data) to comprehensible diagrams and charts. These diagrams and charts can enhance the ability of people to explore, capture essential information promptly and come up with more application solutions, as well. In this section, we provide useful visualisation techniques supporting content-based information, which are correlated and hierarchical visualisation.

3.1.1 Visualising Correlation

Correlation visualisations are very helpful to determine whether there is any relationship between variables.

Matrix Chart This is a compelling visualisation method for analysing, expressing, and understanding the relationship from two to four groups in a matrix format. Numbers or symbols in each cell of the matrix indicate the strength of the relationship between variables of groups. The matrix diagram allows us to compare, match, and search for variables between groups to derive better decisions. For example, the local community and first responders can choose the most effective strategies among possible solutions to behave in and after disasters. There are various types of matrix chart which are L-shaped, T-shaped, Y-shaped, C-shaped, and X-shaped matrices [21], each of them is used depending on how many groups you want to compare. Among these five types of matrix chart, people frequently use the L-shaped matrix chart.

- **L-shaped matrix chart:** among the types of matrix chart, this is the basic, simplest, and most popular matrix diagram to capture a critical relationship between two groups of variables by using a two-dimensional table. The variables

of the first group and second group are placed on the left column and the top row of the table, respectively. The value in the intersection between rows and columns indicates how related between pairs of variables of two groups.

- **T-shaped matrix chart:** it is the combination of two L-shaped matrices to compare one core group with two other groups (e.g., we use the T-shaped matrix to compare two cities that are related to a list of resilient indicators). In this type of matrix chart, we usually represent variables of the core group as columns and variables of the other two groups as rows.
- **Y-shaped matrix chart:** the Y-shaped matrix chart is used for identifying interactions among three groups in a circular manner. Assuming that three groups are a,b,c respectively, we can discover the relationships among these following pairs of groups: (a,b), (b,c), and (a,c).
- **C-shaped matrix chart:** the C in the name of this matrix chart is the abbreviation for *Cube*. This is the extension of the Y-shaped matrix chart. We may use the C-shaped matrix chart for representing and comparing three groups concurrently in a three-dimensional space.
- **X-shaped matrix chart:** this is the extension of the T-shaped matrix, allowing you to conduct the comparison among four different groups. Each group is related to two other groups that are immediately adjacent to it (i.e., two groups, which are opposite, are not related to each other).

Node-Link Diagram Node-link diagram is also known as a network graph or a relationship map. The node-link diagram illuminates interconnections and relationships between a set of entities through the use of nodes and links (or vertices and edges in a graph). Nodes are often displayed as points, dots or circles, but we can use squares, icons, and symbols instead. We connect nodes by links, which are usually straight lines, or curved lines in complicated diagrams. Labels can be used to provide additional information to nodes and links. If there are multiple relationships between two nodes, we may use more than one line. The line colour is useful to show different types of relations as well. Besides, not all of the nodes and links are same in all node-link diagrams. They can have different sizes, shapes, and orientations, depending on the type of node-link diagram that we are using. There are four features of node-link diagram categorised into two groups, which are *weighted and unweighted*, and *directed and undirected*. We can express the weight of node and line by node size and line thickness, respectively. One-way or two-way arrows can represent the direction of relations. Based on four features, there are four significant types of node-link diagram which are: (1) unweighted and undirected, (2) unweighted and directed, (3) weighted and undirected, and (4) weighted and directed node-link diagrams. The node-link diagram is a powerful method to show how entities connect and which entities are more important. To derive useful insights from a node-link diagram, we should focus on nodes with many connections to discover clusters, central nodes, trivial correlations, and connectivity patterns. However, the use of so many nodes and links will reduce the legibility, and we are in trouble to explore new information. In this case, we should focus on particular sub-diagrams or eliminate some nodes and links with

pre-defined rules. Also, we can leverage two traditional approaches which are force-directed layouts and edge bundling [12] to construct a better node-link diagram with many nodes and connections. There also have other similar diagrams which are the chord diagram, the Sankey diagram, and the arc diagram. We may use these types of diagram instead if your data set does not work with the node-link diagram well.

Word Cloud Word cloud given a textual corpus, we may have to tackle with a hundred thousand different words that seem to have no clue for exploring their correlation. Word cloud, otherwise called text cloud or tag cloud, may be useful in this situation. This is a visualisation method to show a collection of words appearing in different sizes. The larger size of a word (or sometimes the bold of text), the higher frequency this word appears in the document; with the assumption that words with higher frequency are more important. The colours in the word cloud are often used for aesthetic or classifying types of words. We can use word cloud visualisation to identify wording, recognise semantic similarity, analyse and understand users' sentiment, discover underlying patterns, and optimise search engines. Remembering to pre-process your textual data before visualising with word cloud; otherwise, we may be unable to derive useful information. There are different pre-processing tasks which we can do at the first step to remove noise, some of which are lowercasing, lemmatisation, stop-words deletion, and word normalisation [18]. Word cloud visualisation, in general, are effortless, fast, attractive, and easy to understand. Nevertheless, there are several situations that we should not use the word cloud. That is when the frequency of words is not adequate different, when words in a document are not variable enough, and when our textual data is too noisy. Besides, long words may get over-attention than short ones. The designs of the word cloud are also vital points to consider. The use of so many fonts (or tangled fonts) and colours, the cramped distance between words, the large number and messy direction of words, and the complicated shape of a word cloud can affect audiences in recognising importance information as well. We must guarantee that we understand our data and design objectives comprehensively before deciding to use word cloud visualisation.

3.1.2 Visualising Hierarchy

Hierarchical data visualisations are used for displaying multiple groups of data in organisational order. In this section, we present two popular hierarchical data visualisation techniques that are tree diagram and sunburst diagram.

Tree Diagram The tree diagram is a method of visualising the organisational hierarchy of data in a tree-like structure without containing any cycle. Typically a tree diagram consists of nodes (including root nodes, nodes, and leaf nodes) and links; every node has at least one relationship. The root node is the highest one with no parent nodes. From a root node, there are many branches leads to nodes that are connected by links. Each node has either parent or child nodes. We call nodes at the same level and share the same parent node as sibling nodes. The leaf

nodes, which are the ending points of a tree diagram, do not have any further child nodes. We usually use rectangles or circles as nodes, straight or elbowed lines as links, and descriptive text inside or around nodes. With the recursively defined property, any subtree is also a completed tree diagram. The root node usually has an overall meaning. From top nodes to leaf nodes, the content becomes more specific and detailed. Tree diagram visualisation appears in various applications, from computer science, biology, and mathematics to business, information management, and emergency management. For example, we can build a tree diagram for quick disaster responses with leaf nodes are specific actions. An alternation of tree diagram is the treemap that shows hierarchical variables by rectangle areas. In the treemap, all the categories are represented as rectangle areas. The rectangle area of a parent category will contain all its child categories inside. Comparing to the tree diagram, we can easier recognise the differences between groups in the treemap.

Sunburst Diagram This is a multi-level pie chart through the use of a series of rings. The sunburst diagram will become the doughnut chart if there are no hierarchical relationships. The root node is at the centre, child nodes moving outward from there and placing on top of each other, and leaf nodes stay at the outermost of the circle. Each ring in the sunburst diagram represents for a level of the hierarchy. The slices of every ring are divided based on hierarchical relationships with their parents. The angle of slices having the same parent can be either equal or proportion to their parent/the whole diagram slices. Different colours can categorise different types of variables or hierarchical levels. Utilising sunburst diagram helps us understand the hierarchical relationship between outer and inner rings. Along with hierarchical information, sunburst diagram can be sufficient to show the part-to-whole relationships, i.e., between a variable with either its parent variable or the whole chart. We should not use sunburst diagram if there are so many slices in each ring because this can skew our perception. Besides, labels will be ineffective and useless if the space of slices is too cramped. Instead of using a label for every slice, we may use for inner ones only (e.g., at the first or second level) because they will have a larger area.

3.2 Visualising Geospatial Information

Visualising geospatial information is one of the earliest approaches in the history of information visualisation. This is the process of representing objects, elements, and events using their location. The location is a diverse range of areas; from a small place such as an office, a building, or a street; to a large region like a city, a nation, or a continent. On the opposite with static information, as we mentioned, the location can also be dynamic. It means the movement from a location to the others, referred to as spatial interactions [9, 24]. For example, the movement of a storm, the expansion of a fire forest, or the spread of disease. Geospatial data can be captured either by humans (e.g., geologists, land surveyors, photographers, or

polices) or by machines (e.g., sensors or GPS-enabled smartphones). Leveraging geospatial information can bring us a lot of benefits in various applications [15].

3.2.1 Geospatial Data Types

After collecting, we can represent and store geospatial information in various forms including vector, raster, compressed raster, geographic database, relational database, light detection and ranging (LiDAR), computer-assisted drafting (CAD), elevation, web, multi-temporal, cartographic, 3D, and interchange file formats.² For the sake of simplicity, we only provide the most commonly used forms as follows.

Vector Vector data is beneficial for modelling discrete objects such as streets, rivers or buildings. This is the basic type of data that people always start thinking of when they face with spatial data. It is the combination of vertices and paths for representing location and shape of geographic features, by three geometric shapes which are points, lines, and polygons (or areas). Ubiquitous online mapping application, some of which are Google Maps and Open Street Maps, represent their geospatial data using this format.

- **Vector point:** each point has its location using (X, Y) coordinate or longitude and latitude values. Each vector point can describe information itself. In the topic of emergency, we can describe a traffic accident, a destructed building, or a victim location as a point.
- **Vector line:** is the connection by at least two points with or without direction. A vector line starts and ends with nodes and changes its directions through vertices. Information can be attached to a specific point, node, or even an entire line. Examples of emergency events which are represented well by vector lines are the shifting of a storm and the movement of rescuers.
- **Vector polygon:** polygons join a set of points which share the same starting and ending coordinates. We usually place information of a polygon in its centre, be independent of the shape of this polygon. A flooding risk, a forest fire, or a dam breach are well modelled by vector polygon.

Raster In contrast with vector data, we use raster data to store and represent connected objects such as population distribution and temperature of an area. Raster data consists of a set regular grid array, pixels, or cells. Each cell in the raster has a coordinate value that means the position of the centre of this cell. The coordinate value depends on its dimension, in 2-D is width and height, in 3-D is width, height, and depth. The shape of a cell can be square or rectangular; however, it is usually represented as a square. The continuous rasters have associated values gradually changing in a defined manner. We find it essential to take into account the resolution of a raster (i.e., the size of a cell). The resolution can express how large area a raster

²<https://gisgeography.com/gis-formats>.

can cover. If the resolution is too high, a cell will include a small area; therefore, the information in a cell is less. If we zoom in to the map utmostly, we can observe separated cells, each of them brings a particular colour. This concept is especially crucial towards geospatial visualisation. By selecting the optimal size of a cell, we can effectively and efficiently express necessary information while minimising file size.

Web Files As we are living in the area of internet and big data, publishing geographic features over the internet is an essential requirement. The GeoJSON and the GeoRSS file formats were built to store and display geographic features over the internet effectively. GeoJSON is designed based on the JSON (JavaScript Object Notation) standard format that contains both geospatial and attribute data. The GeoJSON filename usually ends with *.geojson*. This web file data has been widened used by popular services and systems (e.g., the QGIS, the ArcGIS, the Tableau, and the Spotzi). Meanwhile, GeoRSS is developed based on Geography Markup Language (GML) for describing and pinpointing geographic information on Internet content. These two formats can well describe complex natural features such as canyons, lakes and rivers to human-made creations such as buildings, universities, and cities. Besides that, there exist cloud-based platforms (e.g., Esri ArcGIS Online Web Services) that allow individuals and organisations to publish their contents in shareable and recoverable environments.

3D 3D data is the extension of 2D data by adding Z-aspect to the dimension for creating a triple coordinate (X, Y, Z); therefore, it is similar with vector data and raster data in term of concept. Z-value can be either a tangible value (e.g., geological depth) or an intangible value (e.g., the suitability of a place, the level of pollution, or concentration values). A 3D geospatial surface can be represented by a connected triangle. There are two basic types of 3D data which are feature data and surface data.

- **3D feature:** represents 3D geospatial information for discrete objects.
- **3D surface:** expresses continuous phenomena by having height values over a specific area.

Overall, types of geospatial data in which we select are depend on our input data, expected output, and targeted audiences. Each type is not higher-level than the others, but each can maximise efficiency when it is used in the right demand and context. If the output is close to traditional cartographic representations, then vector data would be appropriate. Meanwhile, raster data is more suitable in term of representing a surface, physical phenomena, or mathematical context.

3.2.2 Techniques

Geospatial data convey a physical context, mostly in 2D space, like geographical maps or floor plans. The primary form of visualising geospatial data is mapping. In *choropleth* maps, colour encoding is used to add represent one data attribute.

Cartograms aim to encode the attribute value with the size of regions by distorting the underlying physical space. Tile grid maps reduce each geospatial area to a uniform size and shape (e.g., a square) so that the colour coded data are easier to observe and compare, and they arrange the tiles to approximate the neighbour relations between physical locations. Grid maps also make a selection of smaller areas (such as small cities or states) easier. Contour (isopleth) maps connect areas with similar measurements and colour each one separately. Network maps aim to show network connectivity between locations, such as flights to/from many regions of the world. Spatial data can also be presented with a non-geospatial emphasis (e.g., as a hierarchy of continents, countries, and cities). Maps are commonly combined with other visualisations. Based on types of geospatial data, i.e., data which contains geo-referencing, there are popular visualisation techniques to show and leverage full advantage of geospatial information as follows.

Cartogram A cartogram describes information by using the forms of geographical regions. Nevertheless, we can distort or modify the actual shapes of geographic areas (i.e., expanding or reducing the actual size of the geographical regions) to best fit with our data. Although a cartogram is not correct in term of geographic size, it still preserves the spatial relationships of objects. There are several popular types of cartogram which are non-contiguous cartogram, contiguous cartogram, and Dorling cartogram.³

- **Non-contiguous cartogram:** this is the simplest one among different types of the cartogram. Non-contiguous cartogram does not conserve the connectivity of adjacent objects. This detachment allows geographical objects to expand or shrink their size without distorting their natural shape. There are two types of non-contiguous cartogram which are non-overlapping (for avoiding overlapping but affecting the distance of objects) and overlapping (for maintaining centroid coordinates of objects). Among the two types, people usually prefer to use non-overlapping because we can observe objects comprehensively.
- **Contiguous cartogram:** comparing to non-contiguous, the contiguous cartogram is more complicated to construct because we must preserve the topology between geographical objects. This type of cartogram immensely distorts the shape of geographic regions to represent attribute value associated with this region. The higher value, the more distorted the size of geographical areas will be affected.
- **Dorling cartogram:** in Dorling cartogram, we represent the geographical objects as uniform non-overlapping shape circles, with appropriate positions and suitable size that can maintain their original topology effectively. The Dorling cartogram can preserve not only the topology but also the shape, and centroid of geographical objects. Similar to Dorling cartogram is the Demers cartogram in which circles are replaced by squares to decrease the distance between terrestrial objects.

³[http://www.ncgia.ucsb.edu/projects/Cartogram\\$_\\$Central/types.html](http://www.ncgia.ucsb.edu/projects/Cartogram$_$Central/types.html).

There are other types of a cartogram, including the hexmaps cartogram, the tilegrams cartogram, and the distance cartogram as well; however, we only describe a more detailed for popular types as above mentioned. Every kind of cartogram can represent geographical objects from a different perspective. Although we can gain many advantages by visualisation geospatial information as a cartogram, it also has limitations in shape recognition and area magnitude estimation. In [6], authors suggested overcoming these problems by (1) representing the shapes of the enumeration units as irregular polygons and (2) “at least one square legend symbol should be used at the lower end of the data range. It is best to provide three squares in the legend, one at the low end, one at the middle, and one at the high end of the data range”.

Flow Map A flow map is a combination between a flow chart and a map to represent the movement of objects between different areas and geographically express the distribution. In a flow map, the arrow depicts the direction, and the thickness of lines represents for the magnitude or amount of phenomena of objects. The use of flow lines is similar to other graduated symbols on other thematic maps [2].

One of the benefits of a flow map is that it enables users to easily recognise the differences in magnitude or amount of a wide diversity of items across areas without many map-clutters [20]. This allows cartographers and Geographic Information Systems (GIS) analysts to see the majority of the movement of objects, and then they can discover implicit patterns. Almost flow maps are created with vector, instead of raster data, because the changes of objects usually shown as lines. In vector-based flow maps, the vectors are points or lines that represent information about the direction and magnitude of items which are transferring. There are three different types of flow maps⁴ as follows.

- **Radial flow map:** this type of flow map represents the relationships between one source towards many destinations. It uses different lines radiating out from a starting point to express the movements.
- **Network flow map:** network flow map depicts the quantity of flow in an existing network; hence, this type of flow map can effectively show transportation systems or communication networks.
- **Distributive flow map:** similar with radial flow map, the distributive flow map show relationships between a single source and many destinations; however, this map often has a large, single line from one source and this line is divided into different parts when reaching its destination.

We can use a flow map for showing dynamic geospatial information of people, weather phenomena, and other living things (e.g., the migration of people, the movement of storms across space, or river flow).

⁴<https://www.gislounge.com/overview-flow-mapping/>.

Density Map Density map is sufficient for visualising the density differences between areas. To represent how different between objects, people usually use colour scale either with a linear mapping [13] or with a non-linear mapping [10]. This type of visualisation brings many advantages towards big dynamic data because we can have valuable insight at the application level. We can create a density map by using either point or line data. The way we choose radius value can affect the representation of the density map. With larger values of the radius value, the density map is more generalised. On the opposite, the smaller values of the radius can show information more detail. There are different methods for calculating magnitude, which is point density and line density. These two interpolation methods can provide quantitative values to represent the concentration of points and poly-lines.

- **Point density:** this method calculates the magnitude-per-unit area of point features that fall within a neighbourhood around each cell. We consider only points that fall within the neighbourhood, if there is not exist any point, we understand as no data.
- **Line density:** this method is similar to point density; nevertheless, it calculates based on poly-line features instead of point features. To obtain an appropriate value of density, we need to select a suitable area unit scale factor (e.g., square kilometres and square meters). The larger the value of area unit scale factor, the more significant value of density we have.

The magnitude of point or line at every sample location is spread through the study area. Density map can determine which locations are greater or fewer numbers of data points or lines. We can utilise density map for visualising different information, some of which are population, urban [25], bomb, crisis [7], and even unhealthy behavior of men [22]. One practical example is the use of a density map for tornado monitoring and analysis in the United States from 1950 to 2017.⁵ This visualisation is extremely useful for meteorologists to anticipate weather and to estimate where a tornado may move to have a good preparedness, response, and mitigation of damages [5]. Density map is more effective in case there are many data points (or data lines) in a small geographic area.

With the development of social networking services and smart devices, there emerge many applications of density mapping using geo-located Twitter data to discover hidden patterns of social events. In [19], authors collect geo-located tweets in 1×1 km grid cells over 2 months in Indonesia. By applying random forests-based census dis-aggregation method to the geospatial data and density mapping, we can comprehend population distributions at a particular time. Governments can take advance of this work for having adjustments promptly to increase the citizen's quality of life. In another work, authors create a density map of Manhattan urban using the data of billion taxi trips for over 7 years [30]. This research is useful for detecting and understanding civic events given specific locations.

⁵<http://maxfelsenstein.com/gis-maps/tornado-density-map-1950-2017/>.

3.3 *Visualising Temporal Information*

Time is a distinctive dimension in real world. It gradually moves forward without our control. But, we often record it as a moment or interval. We can represent it in different ways, for instance year, month, day, hour, and minute. As we calculate time based on cyclic (day/night) incidents in environment, our representations are cyclic too. For example, May follows April and so on. The cyclic way of representation can be described by circular visual encodings, such as the typical clock. Time series data is data that is collected at different points in time, such as stock market or weather data. Data points in time series are gathered at adjacent time periods and has potential for correlation between observations. The emphasis of the analysis is on exploring temporal trends and anomalies, probing for precise patterns, or prediction. The statistical attributes of time series data often violate the notions of standard statistical methods. Thus, analysing time series data requires a unique set of tools and techniques, mutually known as time series analysis. Temporal analysis is comprehending the sequences of events. In our day-to-day life we analyse event sequences. In the analysis of event sequences, uncovering the most common patterns, perceiving unique patterns, searching for certain sequences, or knowing what directs to certain kind of incidents is important (e.g., what situations lead to a unrest during the football match). In this section, we mention different methods for visualising temporal information effectively, which are a line graph, area chart, polar area diagram, and the Gantt chart.

Line Graph We use a line graph to show how quantitative values of something changed over a continuous interval or period. A line graph is constructed by connecting individual data points in the Cartesian coordinate system, consisting of a horizontal x-axis to show a timescale or a sequence of intervals (e.g., hours, days, weeks, or months) and a vertical y-axis to show quantitative values of data. The values on the x-axis are independent because it remains unaffected by other values. On the other hand, the values on the y-axis are dependent because a y-value must correspond to an x-value. Hence, we can call x-axis as independent axis and y-axis as the dependent axis. At a particular time, each x-value only has one y-value associated with it; nevertheless, different x-values may have the same y-value (e.g., a country only has one value of the population at a specific time). Typically, the y-value is positive, but we can express negative values under the x-axis as well. The label in each axis should be selected and divided into suitable units according to different data sets; for example, the x-axis would represent the time measured in years for the population of a country and in months, days, or hours for average temperature. The line graph may lose its clarity if there are so many points in a tiny area.

Besides, we can draw more than one line in the same chart to discover how different between variables, for example, we can compare the temperature, the number of storms, or the population between countries. In this circumstance, the lines should have different colours, shapes, or patterns. However, we should

avoid drawing too many lines in a graph because the line chart becomes denser, challenging to see, and impossible to get insights from. A good number of lines to represent information effectively and efficiently is under five.

Area Chart Area chart is the mixing between the line graph and bar chart. Similar to the line graph, an area chart can show the changes or patterns of quantitative values over time. For example, we can comprehend seasonal peaks in the area chart as a periodic pattern. The x-axis typically represents time, and the y-axis is for another variable that depends on the x-axis. We connect discrete data points by straight lines or smooth curves. An area chart which connects points by a fitted curve is defined as the spline-area chart. The difference with a line graph is that the space between line and x-axis in an area chart is filled up with specific colours, shadings, or patterns. By filling the area under the curve, we can observe the trend from data more apparent. This is also a limitation of the area chart. Comparing to the line graph, we find it harder to add data labels because of less available space.

To use or not to use the area chart will depend on our data set. An area chart can work best if we meet the following requirements: (1) the comparison between values towards the total value is essential, (2) the difference between values are adequate, and (3) the data set is represented over a long period with many values of the time. Though the basic area chart is good at showing how values develop over time, it will not be effective when tackling with multiple attributes. In addition to the primary area chart, there are extended versions as follows.

- **Stacked area chart:** this is the extension of a basic area chart, which can deal with several groups on the same graph by placing each group on top of another one. When only one attribute is presented, the stacked area chart becomes the basic area chart. The stacked area chart is useful for figuring out how the total value is distributed to groups. We may use the stacked area chart for negative numbers, but typically people use positive values instead. By normalising values at each timestamp to the range between 0 and 100%, we can draw a percent stacked area chart. It is precisely the same thing, but the y-values are always on the 0–100% scale. The percent stacked area chart can represent the performance of each segment concerning the total, but we will be unable to find information about the trend of the total.
- **Stream graph:** a stream graph is a modification of the stacked area chart to display high-volume datasets along a different central horizontal axis. As the name, this type of visualisation represents the changes of values of several groups over time by resembling a themed river with the use of flowing and organic shapes; therefore, there is no fixed, straight corner, axis, or angle as in the stacked area chart. This makes the stream graph more interesting and entices users to see the graph. We should use the stream graph when groups are possible to start and finish at different time points. In a stream graph, the height of stream shapes is corresponding to the values of groups. The vertical dimension does not imply positive or negative values. There may not exist even a y-axis to use as a reference

in many stream graphs. Hence, we should focus on the general view of the stream graph rather than spending much time to concentrate on a slice at a particular time point.

Polar Area Diagram Polar area diagram can display data occurring in cycles (e.g., months, years, and seasons) cleanly and effectively. It is also known as a rose chart or a coxcomb chart. A polar area diagram is the combination of a radar chart and a stacked column chart. It is similar to a traditional pie chart; however, sectors have the same size of angles. The distance that extends out from the centre of the circle representing the value of each variable. We can express multiple groups by continuously stacking each on the others in a sector of the pie. To create an excellent and correct polar area diagram, we should keep in mind these experience carefully: (1) each the data set has to follow the part-to-whole relationship (i.e., the sum of values in a group is 100%), (2) selecting appropriate colours (e.g., the colours between groups should be disparate, using darker colours for groups having higher value), (3) we should avoid using too many numbers of sectors to ensure the visibility and legibility, (4) a functional polar area diagram should have the area of sector corresponding to the value of the variable that is representing, and (5) adding data labels may not useful towards polar area chart because of the limited space. Comparing to the traditional pie chart, the polar area diagram is a little harder-to-understand; however, it can display multiple data sets in the same graph, instead of using a series of traditional pie charts.

Gantt Chart The Gantt chart is an organisational visualisation tool allowing you to keep track of and update statuses of various activities/tasks/events to guarantee that we can complete them entirely and punctually [17]. This is extremely useful for emergency management in any size. Nowadays, the Gantt charts are widely used for every management activities due to its efficiency and simplicity. There are two primary components of a Gantt chart that are activity and time duration. Very simple, we put on the left of the Gantt chart a list of all activities and on the top a timescale. Based on the original bar chart, the Gantt chart uses the horizontal axis to display the progress of all events that need to be accomplished, with each one is a bar. The position and length of horizontal bars represent for the starting time, ending time, and duration of activities. We can use different colours for categorising different types of events. In Gantt charts, we may display additional information such as the contributors, milestones, or dependency relationships (i.e., how an action relates to the others by connecting arrows). The current status in the Gantt chart can represent by using different colours or shadings, filling in partially, or plotting a vertical line of the present day. Because activities can run in parallel; it is crucial to control and ensure that there are not so many activities overlapping at a particular time point. With the Gantt chart, we can recognise overlapping events efficiently and promptly. There are two popular types of Gantt chart which are the progress Gantt charts and the linked Gantt charts.

- **Progress Gantt charts:** in progress Gantt chart, we shade horizontal bars until the positions where the tasks have completed (i.e., a job that has completed 50%

should be 50% shaded). We can use a vertical line as the current time to monitor the progress of all task. If parts of horizontal bars of all tasks at the left side of the vertical line shaded, everything is processing punctually. This type of Gantt chart provides an excellent visual representation of the progress of tasks to determine which ones should be taken into account soon. Being unable to manipulate workflow is one disadvantage of the progress Gantt chart because it does not display the dependencies between tasks explicitly.

- **Linked Gantt charts:** the linked Gantt charts represents the connection between tasks by lines. Nevertheless, linked Gantt charts can be very cluttered and lose its clarity if a project has many tasks that are strongly dependent on the others. This is also a limitation of the linked Gantt chart in expressing complex interdependencies between tasks, which may usually happen in large emergency projects.

Temporal information visualisations are the simplest and quickest techniques to represent important information about features and attributes regarding the duration of time [4]. The visualisation can enable and enhance the ability to discover different social events and phenomena in term of temporal data (e.g., crime trends, disaster patterns, and temperature changes). Temporal data is simply data that contains temporal information, usually having a starting and a finishing time. There may have overlap between events. We create temporal data in almost activities, for records, management, and presentations. To obtain temporal data, we can collect from different sources manually and automatically; some sources are mass media source, social networking services [11], observational sensors and simulation models. Comparing to geospatial information, we can obtain temporal information more efficiently and seamlessly. With the personally-identifiable nature of the data, we can not access geospatial information in several situations.

3.4 Dashboards

Dashboards can increase situational awareness so that problems can be noticed and solved early and better decisions can be made with up-to-date information [26]. A dashboard exhibits a smart overview of your most crucial information. Dashboards allow you to display your data, gain a new bird's-eye view on emergency response or management and share information with your team. We define dashboard as:

A dashboard is a glimpse of geographic information that helps you observe incidents or activities. Dashboards are designed to exhibit multiple visualisations that work together on a single screen. They offer a comprehensive and engaging view of collected data, to provide key insights for immediate decision making.

Dashboards are a visual design pattern that integrates components (dashlets) that can be interconnected or independent. The dashboard is powered by data from different sources via a pivot (service), with which the components (dashlets) exchange synchronously or asynchronously, delayed or in real-time. Interactivity

can be expound at both levels-individual dashlet and dashboard. The latter is based on the dependencies defined between the components. The data that you visualise can be processed both in the pivot and/or during visualisation. There are also dashboard implementations as stand-alone systems in the organisation's information infrastructure. For the effective application of a dashboard in infrastructure, whether as a standalone element or as part of a specific software solution, the following aspects should be taken into account:

- **Data adaptation:** very rarely a dashboard exhibits data from a single source. Adaptation to the different formats is needed for data to be processed for visualisation.
- **Adaptation of the visualisation:** the variety of visuals often makes it hard to choose the right presentation. Adaptation implies evaluating variants that are at hand for visualisation, which further reflects on the data format.
- **Navigation adaptation:** interplay in a dashboard can be divided into two groups-the first in terms of the data defining the set of actions available with the dashboard at a time, and the second in view of the way in which it can be executed. Dashboard navigation is an rigorous process in terms of computing and communication through dashboard pivot down to the data sources.
- **Supporting context:** the context of the dashboard is determined by the selected sources, the current users, the selected indicators, and the navigation actions carried out so far. Upholding a complex context affects cognitive integration. Through it in the different components (dashlets) on a dashboard, the diverse aspects of the related data are interpreted and visualised.
- **Managing complexity:** to efficiently design and develop a dashboard in management processes, it is essential to manage the complexity that has many dimensions such as semantic transparency, perceptual discrimination, cognitive capability etc.

The utilisation of dashboard holds significant benefit for those with emergency management responsibilities. Emergency managers rely on a wide variety of location-based data to assist their mitigation, preparedness, response, and recovery tasks. However, it can be a ominous task to figure out how to credibly turn the overwhelming flood of raw data into reliable information that can be analysed and shared.

One tool that is helping emergency managers is the use of GIS operational dashboards. GIS technologies are not new; they have facilitated the emergency management community for many years and have become vital elements in modern emergency management practice. GIS takes data that is referenced to an Earth coordinate system and stores, analyses and produces spatial data and information. This information can be collated into a dashboard, which is an interactive visualisation of the information exhibited in a variety of formats, including graphical, maps, or numerical. The strength of a dashboard is that it can convey a vast amount of actionable information on one screen, whether it's a computer or mobile device.

In a very lucid account of how to create a GIS-enabled operational dashboard, creating a dashboard is begun by identifying the desired near-live data feeds, which could come from a location-enabled mobile application, such as a Twitter feed. Connections are then formulated between the GIS and those data feeds. The data from the different sources are then processed and analysed, and the dashboard exhibited on a computer with the ability to access the dashboard on mobile field devices.

The operational dashboards allow emergency managers to comprehend better data from complex, near-real-time data feeds. Also, visualised presentation of information allows emergency managers to gain insights from data that helps them make more informed decisions, take precise actions, and create detailed strategies during a disaster.

4 Research Issues and Readings

The fields of visualisation rely on methods from scientific, geospatial, and information analytics. It benefits from the knowledge out of the field of interaction as well as of cognitive and perceptual science. We continuously see more increasing amounts of data: new sensors, faster-recording methods and decreasing prices for storage capacities in the previous years allow storing huge amounts of data that used to be unthinkable a decade ago. In this section, we will discuss some significant research issues or challenges in the field of data or information visualisation.

Scalability The scalability problem is an enduring challenge for information visualisation. Visual scalability is defined as the capability of visualisation tools to smartly exhibit large data sets in terms of either the number or the dimension of single data elements. Most visualisations handle relatively small data sets but scaling visualisations from millions to billions of records does require cautious coordination of analytic algorithms to filter data or perform fast aggregation, effective visual summary designs, and rapid refreshing of displays. To accommodate a billion records, aggregate markers (which may represent thousands of records) and density plots are valuable. In some cases, the large volume of data can be collected meaningfully into a small number of pixels. For example Google Maps and its visualisation of traffic conditions. A quick glance at the map allows drivers to use a highly aggregated synopsis of the speed of a large number of vehicles and only a few red pixels are enough to decide when to get on the road. Maintaining interactive rates in querying big data sources is a challenge, with a distinct set of methods proposed, such as approximations and compact caching of aggregated query results.

Scalability is a major challenge of data visualisation as it establishes the power to process large datasets by means of computational overhead as well as pertinent rendering techniques. Information visualisation has recently developed numerous techniques to visualise datasets, but only some of them are scalable to the huge data sets used in visualisation. It is the task of visualisation (analytics) to build a

higher-level view of the dataset to acquire insight while maximising the number of additional information simultaneously.

Contrary to the field of scientific visualisation, supercomputers have not been the primary source of data providers for information visualisation. Parallel computing and other high-performance computing techniques have not been used in the field of information visualisation as much as in scientific visualisation. In addition to the standard approach of developing increasingly better ways to scale up sequential computing algorithms, the scalability issue should be studied at different levels—such as the hardware and the high-performance computing levels—as well as that of individual users. Additionally, the challenge of visualising data streams is due to the pattern of the data stream and the necessity to comprehend its contents.

Data Deluge Increasing access to disruptive technologies and the increasing application of sensors are generating massive volumes of data. Such Big Data has huge relevance for emergency management. However, the growing amount of data poses challenges for data management, analysis and verification.

Causality Visual representation, reasoning, and analytics stress the role of information visualisation as the key medium for detecting causality, forming hypotheses, and assessing accessible evidence. The challenge is to derive highly sensitive and selective algorithms that can resolve conflicting evidence and suppress background noises. Complex network analysis and link analysis are crucial in this matter. Due to the exploratory and decision-making nature of such tasks, users need to voluntarily interact with raw data as well as its visualisations to find causality. Methods such as multiple coordinated views will boost the discovery process. Features that facilitate users in detecting what-ifs and test their hypotheses should be given.

Visual Impairment Colour impairment is a common condition that needs to be taken into consideration. For example, red and green are appealing or their intuitive mapping to positive or negative outcomes (also depending on cultural associations); however, users with red-green colour blindness one of the most common forms, would not be able to differentiate such scales distinctly.

Interpretability The ability to recognise and understand the data is one of the key challenges in the visualisation. Creating a visually correct output from raw data and drawing the right conclusions largely depends on the quality of the used data and methods. Several potential quality problems (e.g., data capture errors, noise, outliers, low precision, missing values, coverage errors, double counts) can be hold in the raw data. Also, the preprocessing of data in order to use it for visual analysis shows several possible quality problems. Data can be inherently incomplete or out dated. The challenges are to determine and to minimise these errors on the preprocessing side, and to provide a flexible yet stable design of the visual analytics application to manage with data quality problems. For example, Homeland Security applications, in particular, have to deal with missing values and uncertainty. Suppose a screening program in the context of Homeland Security in a sensitive area. The system should identify potential attackers, but also try to minimise false positives in order to avoid incorrectly targeting innocent commuters. A falsely inserted data

record should not influence the primary manner in which the system observes and analyses different people. Moreover, updated data of a potential attacker might not be available in the database, but the visual monitoring and analysis of patterns should still work even though the records in the database are largely incomplete.

Aesthetics The aim of information visualisation is the insights into data that it provides. It is important to comprehend the representation of insights in order to comprehend how insights and aesthetics interact, and how these two goals could sustain insightful and visually appealing information visualisation. Much of the aesthetics wisdom consists more of heuristics than empirical evidence at the elementary level of perceptual-cognitive activity. Research in this area mostly focuses on graph-theoretical properties and hardly ever involves the semantics associated with the data. Insights should be detected in the data modelling phase of the process. Incorporating aesthetics into information visualisation still remains a challenge.

Semantics Another challenge in the context of information visualisation is to furnish semantics for analysis tasks and decision making visualisation. Semantic meta data extracted from heterogeneous sources may capture associations and complex relationships. Hence, providing approaches to analyse and detect this information is important to visualisation applications. Ontology-driven approaches and systems have allowed new semantic applications in financial services, web services, business intelligence, and national security. Nevertheless, more research is necessary in order to increase capabilities for creating and maintaining large domain ontologies and automatic extraction of semantic meta data, since the integration process between different ontologies to link various datasets is not fully automated yet. To perform more effective analysis of heterogeneous data sources, more advanced methods for the extraction of semantics from heterogeneous data are a key requirement. Thus, research challenges arise from the size of ontologies, content diversity, heterogeneity as well as from computation of complex queries and link analysis over ontology instances and meta data. New ways to resolve semantic heterogeneities to discover complex relationships are crucial.

Evaluation Human-centric evaluation of visualisation techniques can give rise to qualitative and quantitative assessments of their potential quality, with previous researches focus on the effectiveness of basic visual variables. Evaluations can aim to measure and study the size and value of the insights divulged by the employing of exploratory visualisation tools. Diagnostic usability evaluation is a core of user-centred design. Usability studies can be carried out at different phases of the development process to verify that users are able to complete benchmark tasks with adequate speed and accuracy. Resemblances with the technology earlier used by target users may also be possible to verify improvements. Metrics need to address the learnability and utility of the system, in addition to performance and user satisfaction. Usage data logging, user interviews, and surveys can also assist in spotting potential enhancements in visualisation design and developments.

5 Social Media Data Visualisation and Filtering

In the following section an overview of the methods used for social media data visualisation related to emergency events will be given.

5.1 *Sorted List*

A simple way of displaying posts and other news is to arrange them in a simple list. An advantage of this presentation is the easy (re-)arrangement, for example in a chronological, geographical or relevance (based on more complex metrics) order. The posts can also be shown with accompanying pictures (e.g., of authors) or timestamps.

5.2 *Table*

A more structured approach to visualise data in comparison to a list is a table with several columns like frequency, post content, author, etc. In general, a table can be ordered by clicking a column header. The items in the table are then ordered ascending or descending according to the attributes in the selected column.

5.3 *Timeline*

A timeline is a form of visualisation where one axis represents the time. The other axis can represent various types of information like events or posts. The advantage in comparison to a list is a better visualisation of the chronological order as the temporal relationships between events can be easily visualised and understood.

The displayed data can be as simple as points just representing an event (“something took place at this point in time”) but can also be more complex such as a visual representation of what happened at this point in time (i.e., a photo).

5.4 *Structural*

A structural visualisation collects various node-link network layouts to illustrate the structure of a network. It supports the exploration of connectivity in large graph structures. In the case of social media data for example the spreading (reposting) of posts by actors can be shown. Colours can be used to emphasise links or present

additional information. In addition to node-link diagrams there is also a matrix orientated approach that compresses information to get a better use of limited space. This gives them an advantage over node-link diagrams which rapidly clutter when the network grows.

The map is the most often used visualisation method. This is understandable as a map is a medium that can be intuitively understood and is appropriate for visualising geographical information. However most maps contain only basic features such as displaying a marker for every post, maps displaying enhanced data are used less often.

Statistical information is also frequently used as it provides a visual way of understanding data. Humans can use this visual display e.g., to detect anomalies or track the time progression of an event.

Lists and tables are often used since they provide the most basic way of displaying simple information. This makes them suitable for displaying data that cannot or does not need to be visualised like simple content of posts.

Other visualisation methods such as structural or temporal visualisation are not frequently used although they provide great potential for displaying important information. This might be due to the fact that the data required for them is harder to gather and mine.

5.5 Filtering

When trying to visualise too much data at once the visualisation might get cluttered. Therefore filtering options are needed that reduce the information or adjust it dynamically to the user's focus. In dependence to the form of visualisation, filter mechanics might regard time, location, keywords, quality metrics and other attributes of posts.

In time-based filtering, the user can select a start- and end date, a start time and the length of the time window to be displayed. In addition, the user can specify how the data should be replayed (Movie/Loop) and control the replay (Start/Stop). In addition to simple selection elements for the time another visualisation component can be used for defining a time interval where a timeline is used as a filter element and a data visualisation element at the same time.

Moreover, a map can be used for filtering a dataset based on location as the map's viewport defines a (normally rectangular) geographical area. It is also possible to create more precise location filters on the map by defining an area with markers. The selected area can be used to restrict the dataset to events that occurred within the area, allowing queries such as "What is the sentiment of users in this region?" or "Was there a sharp influx of posts in this city?"

Keywords (and especially tags) are an easy to use method to restrict a large dataset to data that is only relevant for a specific topic. Further, information quality metrics can be automatically computed to determine the quality of an information.

Therefore they can be used to restrict the dataset to data that matches certain quality criteria.

Time, location and keywords were frequently used for filtering, presumable because they provide a simple way to filter out messages concerning a specific event out of a larger dataset.

6 Exercises

In this section, we provide some general guidances of how to apply visualisation techniques mentioned above into possible emergencies using a real dataset collected from social data sources.

6.1 Data Set

To provide practical exercises, we use a real dataset collected from Twitter in 2012 about the Hurricane Sandy 2012. Twitter is an useful source to support managing and analysing emergency situations [16]. The Hurricane Sandy (also known as Superstorm Sandy) was the deadliest, strongest, and the most destructive hurricane of the Atlantic hurricane season in 2012. According to the record, it was inflicted almost \$70 billion USD in damage. It was also the second-costliest hurricane towards the United States until now (the first ranking belongs to the hurricanes Harvey and Maria in 2017). Along the path of this storm through eight different countries, at least 233 people were passed away. This dataset collected for the REVEAL project includes around 2,000 tweets with enough textual, geospatial, and temporal information. Besides, each tweet in the dataset was labelled with location entries at the building, street and region levels manually to provide a gold standard for evaluation work. We can obtain this dataset directly on the website of the REVEAL project.⁶ Each tweet has various information encoded as JSON as follows.

Listing 1 A tweet with various information encoded as JSON

```
1 {
2   "favorited":false,
3   "in_reply_to_user_id":null,
4   "contributors":null,
5   "truncated":false,
6   "text":"Seeing the midtown tunnel flooded and with water
7         just flowing down was scary",
8   "created_at":"Tue Oct 30 04:04:41 +0000 2012",
```

⁶<https://revealproject.eu/geoparse-benchmark-open-dataset/>.


```

8   "retweeted":false,
9   "in_reply_to_status_id_str":null,
10  "coordinates":{
11    "type":"Point",
12    "coordinates":[-73.95240391, 40.74147806]
13  },
14  "in_reply_to_user_id_str":null,
15  "entities":{
16    "user_mentions":[],
17    "mentions":[
18      {
19        "indices":[11, 24],
20        "class":"Location",
21        "subclass":"building",
22        "name":"Midtown Tunnel"
23      }
24    ],
25    "hashtags":[],
26    "urls":[]
27  },
28  "in_reply_to_status_id":null,
29  "id_str":"263129269520195585",
30  "place":{
31    "full_name":"Queens, NY",
32    "url":"http://api.twitter.com/1/geo/id/b6ea2e341ba4356f.
33      json",
34    "country":"United States",
35    "place_type":"city",
36    "bounding_box":{
37      "type":"Polygon",
38      "coordinates":[
39        [
40          [-74.042112, 40.489794],
41          [-73.700272, 40.489794],
42          [-73.700272, 40.812242],
43          [-74.042112, 40.812242]
44        ]
45      ],
46    "country_code":"US",
47    "attributes":{},
48    "id":"b6ea2e341ba4356f",
49    "name":"Queens"
50  },
51  "user":{
52    "follow_request_sent":null,
53    "profile_use_background_image":true,
54    "geo_enabled":true,
55    "verified":false,
56    "profile_image_url_https":"https://si0.twimg.com/
57      profile_images/2659914767/511
58      c11c3bd2a216ed19378592d5b35dc_normal.jpeg",
59    "profile_sidebar_fill_color":"f6ffd1",
60    "is_translator":false,

```

```

59     "id":480979191,
60     "profile_text_color":"333333",
61     "followers_count":80,
62     "profile_sidebar_border_color":"fff8ad",
63     "id_str":"480979191",
64     "default_profile_image":false,
65     "location":"",
66     "utc_offset":-14400,
67     "statuses_count":3764,
68     "description":"24.Virgo.NewYorker.Queens",
69     "friends_count":112,
70     "profile_link_color":"0099CC",
71     "profile_image_url":"http://a0.twimg.com/profile_images
    /2659914767/511c11c3bd2a216ed19378592d5b35dc_normal.
    jpeg",
72     "notifications":null,
73     "profile_background_image_url_https":"https://si0.twimg.
    com/images/themes/theme19/bg.gif",
74     "profile_background_color":"FFF04D",
75     "profile_background_image_url":"http://a0.twimg.com/
    images/themes/theme19/bg.gif",
76     "screen_name":"xnancyi",
77     "lang":"en",
78     "profile_background_tile":false,
79     "favourites_count":96,
80     "name":"Nancy",
81     "url":null,
82     "created_at":"Thu Feb 02 04:39:37 +0000 2012",
83     "contributors_enabled":false,
84     "time_zone":"Atlantic Time (Canada)",
85     "protected":false,
86     "default_profile":false,
87     "following":null,
88     "listed_count":0
89   },
90   "in_reply_to_screen_name":null,
91   "retweet_count":0,
92   "geo":{
93     "type":"Point",
94     "coordinates":[40.74147806, -73.95240391]
95   },
96   "id":263129269520195585,
97   "source":"<a href=\"http://twitter.com/download/iphone\"
    rel=\"nofollow\">Twitter for iPhone</a>"
98 }

```

6.2 Visualisation Tool

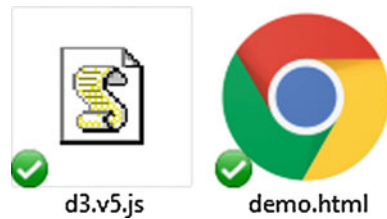
Visualisation tools can organise data in a meaningful way that lowers the cognitive and analytical effort required to make sense of the data and make data-driven decisions. Users can scan, recognise, understand, and recall visually structured representations more rapidly than they can process non-structured representations. The science of visualisation draws on multiple fields such as perceptual psychology, statistics, and graphic design to present information, and on advances in rapid processing and dynamic displays to design user interfaces that permit robust interactive visual analysis.

In this section, we use the D3 library as the visualisation tool. The D3 is a very popular JavaScript (JS) library to give different visualisation techniques for various types of information. The reason for choosing D3.js because JS is a light-weight, interpreted, and just-in-time compiled programming language and compatible with every systems and device. Besides, the D3 library gives us the freedom to modify their source code to adapt to our particular requirements.

To start using the D3 library, we need to download the regular version⁷ or the minified version⁸ (i.e., all the white-space were wiped to reduce file size and time loading) from their website. This exercise is writing for the 5.12.0 version; however, it will be mostly similar to the other versions. After downloading and getting the file `d3.v5.min.js`, we should create a folder for containing this JS file. For the sake of simplicity, I will name this folder as *exercises*. The next step is to create an HTML file, name it as *demo.html*, and place into the same folder with the JS file. The folder *exercises* now contains two files as shown in Fig. 1.

Now copying this source code and put into the file *demo.html*. You then only need to open the file *demo.html* with any browser to run the program.

Fig. 1 The folder structure for this exercise



⁷<https://d3js.org/d3.v5.js>.

⁸<https://d3js.org/d3.v5.min.js>.

Listing 2 Template for making visualisation with the D3 library

```

1 <html lang="en">
2   <head>
3     <meta charset="utf-8">
4     <title>Emergency Visualisation Demo</title>
5     <script type="text/javascript" src="d3.v5.js"></script>
6   </head>
7   <body>
8     // The D3 code will come here
9     // .....
10  </body>
11
12 </html>

```

For any D3 script providing in the next section, we should put into the place that we noted as “*The D3 code will come here*”. This D3 JavaScript code will then generate other necessary HTML elements for the visualisation automatically.

Towards beginning users that do not have much experience in the D3 library, it is beneficial to visit their website⁹ and follow their instructions. In the gallery, different authors modified and extended the original D3 library to adapt to their specific visualisation requirements. After we self-determined our visualisation targets, we may search in the collection of examples to pick up the most similar ones.

6.3 A Case Study with the Hurricane Sandy 2012

In this section, we focus on providing different visualisations that can support end-users understanding useful information about Hurricane Sandy 2012 quickly and efficiently. To derive a helpful visualisation, pre-processing and analysing data are beneficial. For the sake of simplicity, we only mention about pre-processing and analysing data briefly here and leave other space for describing visualisation methods. In the previous section, we introduced different visualisation techniques that can deal with content-based, geospatial, and temporal information.

Given no prior knowledge of an event, the word cloud is beneficial that should be conducted at the first step to derive an overview of what has happened. The input of the word cloud visualisation is a set of non-duplicated words extracting from the attribute “text” in the dataset. This set of words should not contain stop words, which are commonly used words (e.g., the, a, an, and in) but empty of meaning. Besides stop word removal, we can do the stemming or lemmatisation to convert words to their base forms. Stemming considers removing the last few characters, but this may lead to incorrect meanings and spelling errors. Otherwise, lemmatisation focuses more on the context and tries to revise words to their meaningful original form. To split a text into words, we may use a unigram, bigram or trigram model or combining

⁹<https://github.com/d3/d3/wiki/Gallery>.

these models. In addition to the processes as mentioned earlier, we also delete hyperlinks, remove all punctuation marks, and change all the letters to lowercase for reducing noise and redundancy. Finally, we obtain 30 highest frequency words in ascending order as follows: *school, news, coastal, high, wind, crazy, good, warning, live, rain, subway, safe, street, hope, bad, manhattan, storm, basement, east, people, house, tsunami, hurricane, city, water, power, sandy, movie, flood, and newyork*. The source code for visualising these words along with their frequency is given as below.

Listing 3 Source codes for the word cloud visualisation

```

1 <div id="visualisation">
2 // The HTML will be generated here
3 </div>
4 <script src="https://cdn.jsdelivr.net/gh/holtzy/D3-graph-
5 gallery@master/LIB/d3.layout.cloud.js"></script>
6 <script type="text/javascript">
7 // List of words
8 var words = [{word: "school", size: "10"}, {word: "news",
9 size: "10"}, {word: "coastal", size: "10"}, {word: "high",
10 size: "10"}, {word: "wind", size: "10"}, {word: "crazy",
11 size: "15"}, {word: "good", size: "15"}, {word: "
12 warning", size: "15"}, {word: "live", size: "15"}, {word:
13 "rain", size: "15"}, {word: "subway", size: "15"}, {word:
14 "safe", size: "15"}, {word: "street", size: "20"}, {
15 word: "hope", size: "20"}, {word: "bad", size: "20"}, {
16 word: "manhattan", size: "20"}, {word: "storm", size: "20
17 "}, {word: "basement", size: "25"}, {word: "east", size:
18 "25"}, {word: "people", size: "30"}, {word: "house", size
19 : "35"}, {word: "tsunami", size: "35"}, {word: "hurricane
20 ", size: "40"}, {word: "city", size: "40"}, {word: "water
21 ", size: "45"}, {word: "power", size: "55"}, {word: "
22 sandy", size: "60"}, {word: "movie", size: "70"}, {word:
23 "flood", size: "80"}, {word: "newyork", size: "90"}]
24
25 // set the dimensions and margins of the graph
26 var m = {top: 10, right: 10, bottom: 10, left: 10},
27 w = 850 - m.left - m.right,
28 h = 850 - m.top - m.bottom;
29
30 // append the svg object to the body of the page
31 var svg = d3.select("#visualisation")
32 .append("svg")
33 .attr("width", w + m.left + m.right)
34 .attr("height", h + m.top + m.bottom)
35 .append("g")
36 .attr("transform",
37 "translate(" + m.left + "," + m.top + ")");
38
39 // Constructs a new cloud layout instance.
40 // An algorithm to find position of words that suits the
41 needs.
42 // The different from one word to the other must be here.
43 var layout = d3.layout.cloud()

```

```

27     .size([w, h])
28     .words(words.map(function(d) {return{text: d.word, size:
29         d.size;}}))
30     .padding(6) //space between words
31     .rotate(function() {return ~~(Math.random() * 2) * 90;})
32     .fontSize(function(d) {return d.size;}) // font size
33     .on("end", draw);
34 layout.start();
35 // Takes the output of 'layout' above and draw the words.
36 // THE SAME from one word to the other can be here.
37 function draw(words) {
38     svg
39     .append("g")
40     .attr("transform", "translate(" + layout.size()[0] / 2
41         + "," + layout.size()[1] / 2 + ")")
42     .selectAll("text")
43     .data(words)
44     .enter().append("text")
45     .style("font-size", function(d) { return d.size; })
46     .style("fill", "#69b3a2")
47     .attr("text-anchor", "middle")
48     .style("font-family", "Impact")
49     .attr("transform", function(d) {
50         return "translate(" + [d.x, d.y] + ")rotate(" + d.
51             rotate + ")";
52     })
53     .text(function(d) { return d.text; });
54 }
55 </script>

```

Figure 2 depicts the word cloud, which is referred from the D3 examples.¹⁰ To use this source code, we need to define an array containing different words and their frequency (e.g., {word: “school”, size: “10”}). Because the frequency of appearance of words extracted from tweets in the Hurricane Sandy 2012 is very high and deviant, the visualisation may lose its clarity (i.e., some words are too big while the others can be too small). To obtain a better appearance of words, we normalise the frequency of words to values lower than 100 following a defined scaling function. We may change the fonts, colours, sizes, and directions of words in the word cloud by modifying the source codes as well. From this visualisation, we can quickly have an insight in where this event might happen (i.e., New York), what is this event (i.e., hurricane and tsunami), and what is effect of the Hurricane Sandy 2012 (i.e., flood, water, city, house, power, and so on).

For the next visualisation, we consider representing geospatial information. With the development of smart devices, there are more tweets posted with geotagged locations; however, the number of geotagged tweets is still in limitation due to users’

¹⁰https://www.d3-graph-gallery.com/graph/wordcloud_size.html.


```

13
14 <script>
15 var sampleData = {"HI": {"color": "#ffffff"}, "AK": {"color": "#
    "#ffffff"}, "FL": {"color": "#fffff0"}, "SC": {"color": "#
    "#ffffff"}, "GA": {"color": "#ffffff"}, "AL": {"color": "#
    "#ffffff"}, "NC": {"color": "#008000"}, "TN": {"color": "#
    "#ffffff"}, "RI": {"color": "#ffffff"}, "CT": {"color": "#
    "#008000"}, "MA": {"color": "#008000"}, "ME": {"color": "#
    "#ffffff"}, "NH": {"color": "#ffffff"}, "VT": {"color": "#
    "#ffffff"}, "NY": {"color": "#ff0000"}, "NJ": {"color": "#
    "#ffb266"}, "PA": {"color": "#fffff0"}, "DE": {"color": "#
    "#ffffff"}, "MD": {"color": "#fffff0"}, "WV": {"color": "#
    "#ffffff"}, "KY": {"color": "#ffffff"}, "OH": {"color": "#
    "#ffff00"}, "MI": {"color": "#ffffff"}, "WY": {"color": "#
    "#ffffff"}, "MT": {"color": "#ffffff"}, "ID": {"color": "#
    "#ffffff"}, "WA": {"color": "#fffff0"}, "DC": {"color": "#
    "#ffffff"}, "TX": {"color": "#fffff0"}, "CA": {"color": "#
    "#ffff00"}, "AZ": {"color": "#ffffff"}, "NV": {"color": "#
    "#ffffff"}, "UT": {"color": "#ffffff"}, "CO": {"color": "#
    "#008000"}, "NM": {"color": "#008000"}, "OR": {"color": "#
    "#ffffff"}, "ND": {"color": "#ffffff"}, "SD": {"color": "#
    "#ffffff"}, "NE": {"color": "#ffffff"}, "IA": {"color": "#
    "#ffffff"}, "MS": {"color": "#ffffff"}, "IN": {"color": "#
    "#ffffff"}, "IL": {"color": "#fffff0"}, "MN": {"color": "#
    "#008000"}, "WI": {"color": "#008000"}, "MO": {"color": "#
    "#ffff00"}, "AR": {"color": "#ffffff"}, "OK": {"color": "#
    "#ffffff"}, "KS": {"color": "#ffffff"}, "LS": {"color": "#
    "#ffffff"}, "VA": {"color": "#fffff0"};

16
17 // Fill colour for above states.
18 uStates.draw("#statesvg", sampleData);
19 d3.select(self.frameElement).style("height", "600px");
20 </script>

```

We use the example of the density map in the D3 gallery for our visualisation.¹¹ For the sake of simplicity, we removed the tooltip function for reducing the length of this source code. Besides, we modified the source code for users to customise the colours of the density map easily. What we need as the input is an array of US states with their respective colours (e.g., “HI”: “color”: “#ffffff”). Here, the colours represent levels of emergencies. We used five different colours as the descending of risks including red (i.e., significant risk to lives exist and significant damage and disruption), orange (i.e., more certain that there is to personal and property safety), yellow (i.e., hazard is possible, be aware of the potential impacts of the hazard), and green (i.e., non-urgent or not serious), and white (i.e., no hazards expected). The colour codes of these five colour are #ff0000, #ffb266, #ffff00, #008000, and #ffffff respectively. Based on extracted locations in the dataset and their frequency of appearance, the red group includes *NY*, the orange group comprises *NJ*, the yellow

¹¹<http://bl.ocks.org/NPashaP/a74faf20b492ad377312>.

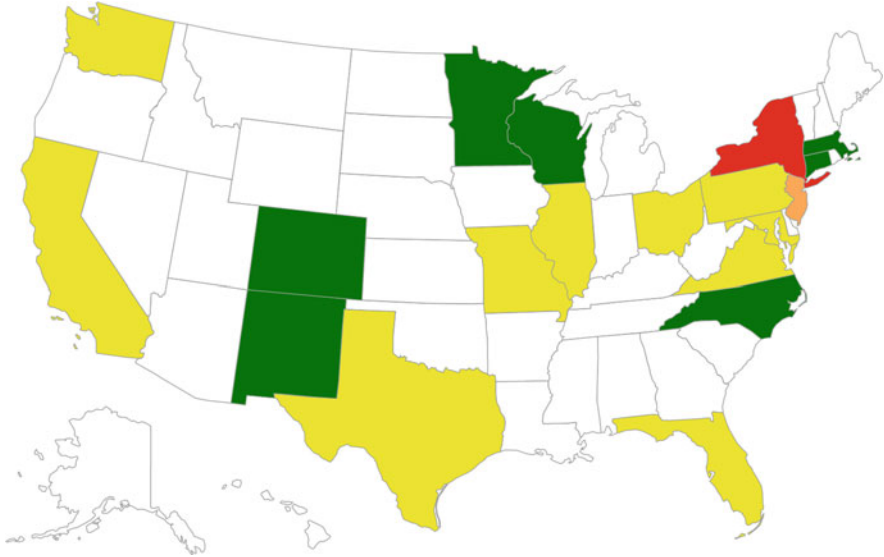


Fig. 3 Density map visualisation of the Hurricane Sandy 2012

group contains *FL, PA, MD, OH, WA, TX, CA, IL, MO, VA*, the green group consists of *NC, CT, MA, CO, NM, MN, WI*, and the rest states belong to the white group.

From the density map in Fig. 3, we may recognise the state under the highest risk was New York, following by New Jersey. However, the dataset may contain noise because the tweets can be posted by not only people in areas of the hurricane but also any person at any state in the US. In case of temporal visualisation, it is not useful towards this dataset because authors collected these tweets in a very short time (i.e., in only 10 minutes from 04:00 to 04:10). In a short duration of time, we find it difficult to express the differences or changes using temporal information. In order to understand the data comprehensively, of course, we need higher-level data analysis with a certain amount of time and effort. Therefore, the former visualisations (i.e., word cloud and density map in this exercise) are very compelling, which can be considered as a roadmap to guide further data analytics.

Acknowledgments The work is funded from the Research Council of Norway (RCN) and the Norwegian Agency for International Cooperation and Quality Enhancement in Higher Education (Diku) grant through INTPART programme.

References

1. Brady, G.P., Stouten, P.F.: Fast prediction and visualization of protein binding pockets with pass. *J. Comput. Aided Mol. Des.* **14**(4), 383–401 (2000)
2. Chang, K.T.: *Introduction to Geographic Information Systems*. McGraw-Hill, Boston (2008)

3. Chen, A.S., Khoury, M., Vamvakeridou-Lyroudia, L., Stewart, D., Wood, M., Savic, D.A., Djordjevic, S.: 3d visualisation tool for improving the resilience to urban and coastal flooding in Torbay, UK. *Procedia Eng.* **212**, 809–815 (2018)
4. Chittaro, L., Combi, C.: Representation of temporal intervals and relations: information visualization aspects and their evaluation. In: *Proceedings of the 8th International Symposium on Temporal Representation and Reasoning (TIME 2001)*, pp. 13–20. IEEE, Piscataway (2001)
5. Deng, Y., Wallace, B., Maassen, D., Werner, J.: A few GIS clarifications on tornado density mapping. *J. Appl. Meteorol. Climatol.* **55**(2), 283–296 (2016)
6. Dent, B.D., Torguson, J.S., Hodler, T.W.: *Cartography: Thematic Map Design*, vol. 5. McGraw-Hill, New York (1999)
7. Fitterer, J., Nelson, T.A., Nathoo, F.: Predictive crime mapping. *Police Pract. Res.* **16**(2), 121–135 (2015)
8. Grinstein, U.M.F.G.G., Wierse, A.: *Information Visualization in Data Mining and Knowledge Discovery*. Morgan Kaufmann, San Francisco (2002)
9. Guo, D.: Flow mapping and multivariate visualization of large spatial interaction data. *IEEE Trans. Vis. Comput. Graph.* **15**(6), 1041–1048 (2009)
10. Herman, I., Marshall, M.S., Melançon, G.: Density functions for visual attributes and effective partitioning in graph visualization. In: *Proceedings of the 6th IEEE Symposium on Information Visualization (InfoVis 2000)*, pp. 49–56. IEEE, Piscataway (2000)
11. Hoang Long, N., Jung, J.J.: Privacy-aware framework for matching online social identities in multiple social networking services. *Cybern. Syst.* **46**(1–2), 69–83 (2015)
12. Holten, D., van Wijk, J.J.: Force-directed edge bundling for graph visualization. *Comput. Graph. Forum* **28**(3), 983–990 (2009)
13. Jerding, D.F., Stasko, J.T.: The information mural: a technique for displaying and navigating large information spaces. *IEEE Trans. Vis. Comput. Graph.* **4**(3), 257–271 (1998)
14. Keim, D., Qu, H., Ma, K.L.: Big-data visualization. *IEEE Comput. Graph. Appl.* **33**(4), 20–21 (2013)
15. Lee, J.G., Kang, M.: Geospatial big data: challenges and opportunities. *Big Data Res.* **2**(2), 74–81 (2015)
16. Martinez-Rojas, M., del Carmen Pardo-Ferreira, M., Rubio-Romero, J.C.: Twitter as a tool for the management and analysis of emergency situations: a systematic literature review. *Int. J. Inf. Manag.* **43**, 196–208 (2018)
17. Maylor, H.: Beyond the Gantt chart: project management moving on. *Eur. Manag. J.* **19**(1), 92–100 (2001)
18. Nguyen, H.L., Jung, J.E.: Statistical approach for figurative sentiment analysis on social networking services: a case study on twitter. *Multimedia Tools Appl.* **76**(6), 8901–8914 (2017)
19. Patel, N.N., Stevens, F.R., Huang, Z., Gaughan, A.E., Elyazar, I., Tatem, A.J.: Improving large area population mapping using geotweet densities. *Trans. GIS* **21**(2), 317–331 (2017)
20. Phan, D., Xiao, L., Yeh, R., Hanrahan, P.: Flow map layout. In: *Proceedings of the 11th IEEE Symposium on Information Visualization (InfoVis 2005)*, pp. 219–224. IEEE, Piscataway (2005)
21. Ryan, T.P.: *Statistical Methods for Quality Improvement*. Wiley, Hoboken (2011)
22. Silva, J.P.: Mapping unhealthy behavior among economically active men using GIS in suburban and rural areas of Sri Lanka. *Asia Pac. J. Public Health* **28**(1 Suppl), 10S–16S (2016)
23. Sivarajah, U., Kamal, M.M., Irani, Z., Weerakkody, V.: Critical analysis of big data challenges and analytical methods. *J. Bus. Res.* **70**, 263–286 (2017)
24. Stock, K., Guesgen, H.: Geospatial reasoning with open data. In: *Automating Open Source Intelligence*, pp. 171–204. Elsevier, Amsterdam (2016)
25. Susaki, J., Kajimoto, M., Kishimoto, M.: Urban density mapping of global megacities from polarimetric SAR images. *Remote Sens. Environ.* **155**, 334–348 (2014)
26. Verbert, K., Govaerts, S., Duval, E., Santos, J.L., Assche, F., Parra, G., Klerkx, J.: Learning dashboards: an overview and future research opportunities. *Pers. Ubiquit. Comput.* **18**(6), 1499–1514 (2014)

27. Vogel, D.R., Dickson, G.W., Lehman, J.A., et al.: Persuasion and the role of visual presentation support: the UM/3M study. Management Information Systems Research Center, School of Management, University of Minnesota (1986)
28. Wang, L., Wang, G., Alexander, C.A.: Big data and visualization: methods, challenges and technology progress. *Digit. Technol.* **1**(1), 33–38 (2015)
29. Yang, H.H., Moody, J.: Data visualization and feature selection: new algorithms for non-gaussian data. In: Proceedings of the 12th International Conference on Advances in Neural Information Processing Systems (NIPS 1999), pp. 687–693 (2000)
30. Zhu, X., Guo, D.: Urban event detection with big data of taxi od trips: a time series decomposition approach. *Trans. GIS* **21**(3), 560–574 (2017)