

TOPSIS-Based Recommender System for Big Data Visualizations

Václav Stehlík*, Mouzhi Ge**

ABSTRACT

Big data analytics can enable effective data interpretation that leads to highquality decision making in organizations. Since the amount of data nowadays has significantly increased in various industries, it thus motivates the research of visualizations for interdisciplinary and collaborative domains. Although there is a large collection of visualization techniques, it is very time-consuming to choose proper visualization techniques for a specific dataset.

This paper therefore aims to analyze the typical and state-of-the-art data visualization techniques for big data. Differing from traditional visualizations such as line chart or bar chart, this paper focuses on reviewing a set of modern visualizations for big data in terms of analyzing their advantages and disadvantages. To facilitate choosing a proper visualization, a recommender model based on TOPSIS is further proposed. In order to validate the proposed model, a prototype of the big data visualization recommender system has been implemented to validate the applicability of the system.

KEYWORDS

Data visualization, big data, data visualization recommendation, TOPSIS

1. Introduction

Big data has received increasing attention in recent years, as organizations and cities are dealing with tremendous amounts of data with high complexity and velocity [1]. These data are fast moving and changing in value, meaning and format. They also can originate from various sources, such as social networks, unstructured data from different devices or raw feeds from sensors. Although there is no unified definition for big data [2], big data can be broadly described as data of substantial size

* Masaryk University, Botanická 68A, 60200 Brno, Czech Republic

** Deggendorf Institute of Technology, Dieter-Görlitz-Platz 1, 94469 Deggendorf
<https://doi.org/10.25929/jair.v1i1.114>

possibly originating from different sources in distinct data format presentations while requiring high data processing speed [3]. One of the prominent approaches of describing big data is the V-model which presents a set of terms starting on letter 'V' that represents various aspects of big data. Initially, three main Vs were introduced by the e-commerce sector by David Laney [4]. Those three Vs were consisting of volume, variety and velocity. Volume of big data refers to the size of data being created from all the sources such as medical data, space images, traffic sensors, satellite imagery, banking transactions, financial market data, etc. [5]. Variety describes heterogeneity of data. Heterogeneity may refer to differences in terms of data types, such as text and images, and data format, or different data sources that generate data such as open data and sensor data. Velocity refers to how quickly data are generated and moved in data storage.

More organizations are focused on interdisciplinary data analytics and intend to use visualization tools to gain insights and leverage it to improve information sharing, increase return of investment, reduce burden of IT, and enable more efficient data driven decision-making [6]. Nowadays, there is a plethora of tools such as Tableau or FusionCharts that offer comprehensive sets of visualizations and help companies to visualize data. However, it requires extensive knowledge and skill sets to select a proper visualization for insightful analysis [7]. Deceptive and inappropriate usages of visualization may hurt the business development, some cases have been well studied such as [8] and [9]. Therefore, if the visualization of data is not selected properly, the previous steps in the big data life cycle such as collecting data, cleansing data, performing analysis, building and validation of models may become invalid.

To help organizations understand and select the proper visualization, the goal of this paper is to first provide a comparative study for various visualizations and then present an approach to recommend the most suitable visualization for specific needs and constraints. Reviewing the visualization is focused on detailed description of advantages, disadvantages, tips and connection to big data characteristics of each visualization. Based on the review of visualization techniques, a set of criteria that may influence the choice of the visualization is selected. Afterwards these criteria are used to build a tool that can help users select a visualization and learn more about how to properly use and take advantage of it.

The contribution of this paper is three-fold. First, the state-of-the-art visualizations are reviewed based on their advantages and disadvantages. This may help organizations, especially non-IT organizations to understand which visualizations can be used. Second, a recommendation model based TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) is proposed to facilitate the visualization selection. This will lead to the efficient selection for interdisciplinary data analytics. Third, a prototype system is implemented with open sources.

The rest of the paper is organized as follows. Section 2 reviews a variety of typical visualizations for big data in terms of summarizing advantages and disadvantages. Based on the reviews, section 3 proposes a TOPSIS-based model that helps users to select the proper visualization according to a specific context. Further, Section 4 implements system prototype to validate the applicability of the proposed model. Finally, Section 5 concludes the paper and outlines limitations as well as future research.

2. Typical Visualizations for Big Data

Data visualization is a representation of information derived from data in a pleasant graphical form that facilitates aspects of communication and insight extraction. This term can be further differentiated into information, scientific and infographic visualization [10]. Scientific visualizations usually have to comply with conventions and expected rules to convey spatial information and closely present real-world physically based objects and models [11]. Information visualization is typically used to represent abstract non-physically based data [11]. Also, infographics usually combine elements of statistical information with data visualization in order to set a narrative or polemic [10].

Visualization can serve as an easily approachable medium of valuable information that allows the user to observe trends, patterns, outliers, relationships, and performance. The most proficient visualizations leverage highlighting, filtering and interactive aspects to remove noise and help the user to focus on the message and important parts [12]. In this section, we will review 10 typical visualization techniques as well as their advantages and disadvantages for big data.

2.1. Treemap

Treemap visualization is used to display large amounts of hierarchical data. It utilizes a root rectangle that is subdivided into smaller rectangles that correspond to the hierarchical structure of the dataset as it can be seen in figure 1. Treemaps are well-known and commonly used on data dashboards, mainly because of their ability to display relationships and hierarchies in data. However, a complex visualization method, Treemap can be hard for beginners to comprehend [13]. Thus, one needs to evaluate the efforts of using treemap visualisation.

Advantages

Treemaps are based on shape volume estimation which is calculated from one or more data factors. Therefore every change in data requires total repaint of the whole visualization for the visible level of hierarchy. Levels that are higher up the hierarchy do not require the action of repainting because they are not visible in the current view [3]. The method allows usage of two measures (data factors) where the first one is size and the second one is color. Size is used for shape volume calculation and color helps with grouping of the shapes. The size measure used for volume calculation must be of computable data type [3]. All of this decreases utilization in regards to data variety.



Figure 1: Treemap depicting stock market from finviz.com [14].

Disadvantages

Velocity is poorly handled by treemaps [3]. Due to its nature, only a snapshot of one moment is displayed and the time component is not present at all. Overview of the development over the course of time would require comparison with a completely new treemap. However, such practises would be cumbersome and the process of making an insight would be less than optimal.

2.2. Sunburst

Sunburst chart, also known as ring chart, multi-level pie chart, belt chart or radial treemap [15]. It is often used to visualize hierarchical data. The origins of this radial space-filling visualization method go back to year 2000 and the research of John Stasko [16]. The radial layout is comprised of a central origin around which rings are stacked in an outward direction [15]. Each ring represents a level of hierarchy. The central origin is labeled as root node. The nodes of the last level, which are located in the most outer parts of the chart, are labeled as leaf nodes or leaves [17]. Rings are sliced up and divided based on their hierarchical relationship to the parent slice. The angle of each slice is either divided equally under its parent node or can be made proportional to a value [15].

Advantages

Advantages of sunburst lie in the ability to show the whole picture of the hierarchy. Sunburst is able to illustrate the different depths of hierarchy and connections between groups. Treemap only provides limited view of the whole hierarchy and requires travelling down or up to view different levels. Another point where sunburst shines is the ability to indicate the depth of each branch. Thus, it is possible to quickly identify outliers and accumulation of values across branches. This results in an additional insight and it also contributes to a better depiction of the overall message.

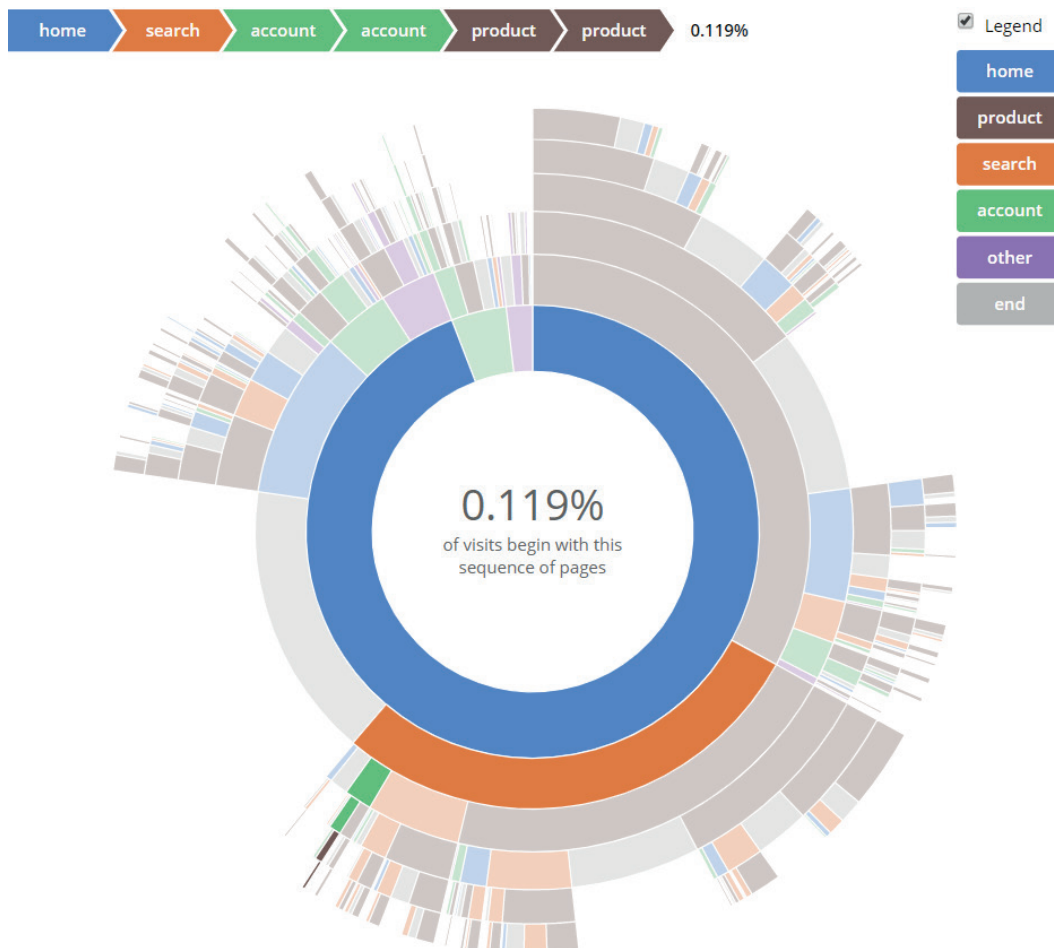


Figure 2: Flow of user actions on website [18].

Disadvantages

One of the ways to describe the data represented on sunburst is to use a legend. Unfortunately, this solution does not scale well with the increasing amount of categories. Therefore, one can employ another commonly used technique for information explanations by labels. Another challenge rises – depending on the size of the node and the length of the label text, it is highly probable that the label will not fit into the visual representation of the node. This may happen if the node is located far away from the origin where high fragmentation of the groups may occur on the left half of the most outer ring. Ignoring this issue will result in cluttering and the chart may even become unreadable.

2.3. Sankey diagram

Sankey diagrams display flows and their quantities in proportion to one another [19]. Directed arrows, called links, connect several entities, called nodes. The width of links indicates the proportional quantity of the flow. The larger the width, the higher the quantity of the flow [20]. Links can be combined together or split through their paths on each stage of a process. Color can be used to divide the diagram into different categories or to show the transition from one node to another [19]. The flow in Sankey diagrams always connects at least two nodes (processes). Thus it shows not only flow values but also information about the structure and distribution of the defined system [21]. Sankey diagrams are suitable choices for displaying movement of energy, money (figure 3), data in life cycle assessments [22] or other critical resources. Further, a Sankey diagram can be used to show the flow of any isolated system process [19].

Advantages

The Sankey diagram is useful for tracking resources inside a system. It allows us to follow the path of resource from the start to end. One can observe the state when entering a certain system and the final output at the end with ability to analyze progress at each node (stage) along the path [24]. It is possible to use interactive aspects and clearly separate single aspect within the system. Therefore, the complex system becomes simplified and optimizations can be made for a given aspect in order to improve efficiency [25]. The width of the individual links and flows provides interesting insights that may reveal extreme values, inconsistencies and other developments. This presents an opportunity that can be used in order to improve efficiency or gain benefits in previously neglected areas [25].

Disadvantages

Sankey diagrams may be hard for people who are not familiarized with data visualizations [25]. Issues may arise when comparing flows and links with similar widths which will complicate orientation [25]. The nodes require appropriate spacing. Otherwise, cluttering and overlapping issues may occur. Therefore, a Sankey diagram should be considered when the space is not an issue [25]. Incorrect placement may also lead to an unnecessary number of crossings between links [24].

2.4. Parallel Coordinates Plot

Parallel coordinates plot or parallel plot is a technique used for plotting and comparison of multi-dimensional numerical data. Parallel coordinates plot consists of several vertical axes where each represents one variable. These axes are placed parallel to each other. Individual data values are depicted as points on axes connected via lines between neighbouring axes and spanning across all dimensions. If placing all the connecting lines of a single object together, we will get a series that is representing this object across all the dimensions. These series of lines can be used to great benefit when it comes to improvements that increase interactivity. Parallel coordinates are helpful for comparing values of different objects across multiple attributes simultaneously.

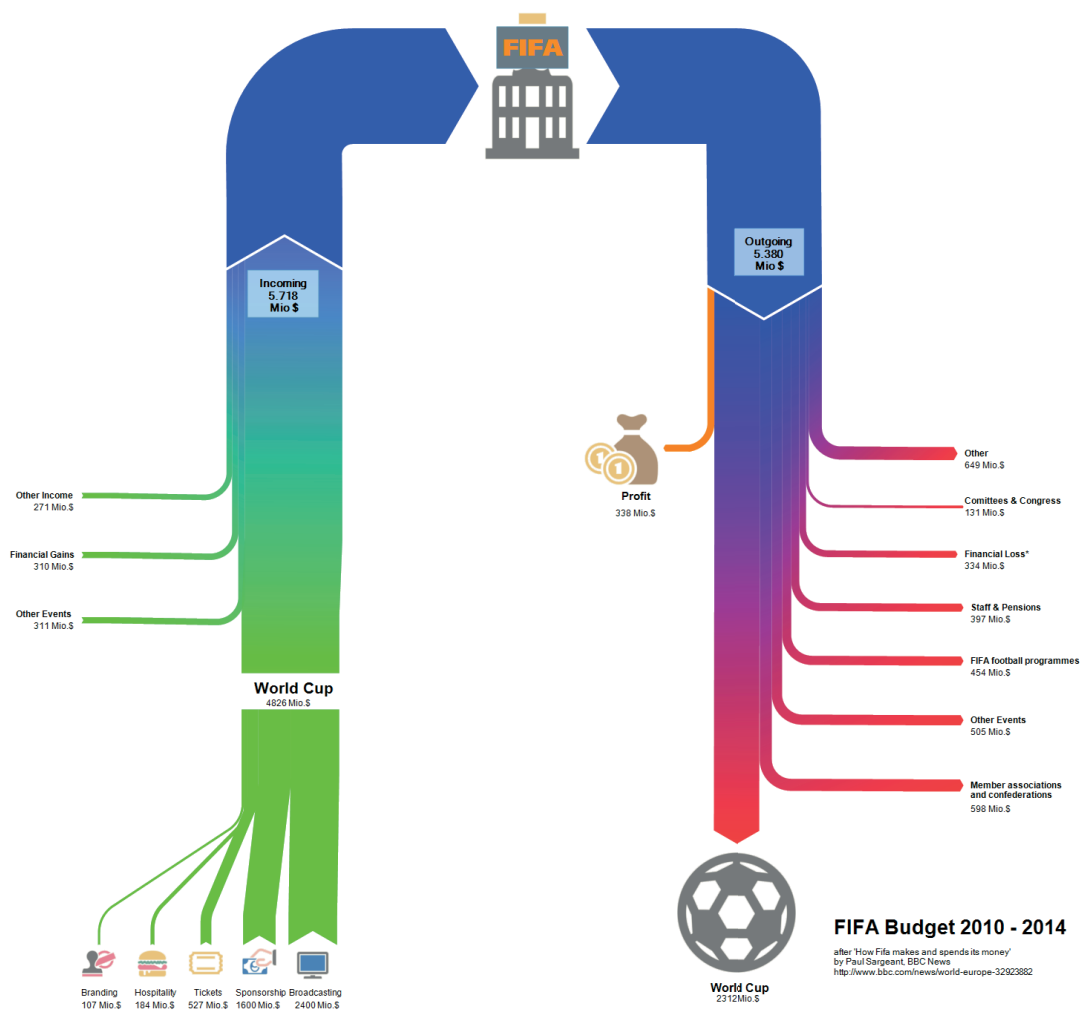


Figure 3: Revenue and outgoings of FIFA from 2010 to 2014 [23].

Advantages

Parallel coordinates plot has several variations that can be suited to specific needs and provide better insights. For example, regarding scaling [27], plot can use one scale for all dimensions which can be normalized or transformed to the desired range. It is also possible to have different scales on each axis as each dimension usually works with a different unit of measurement. Also, the axes can be rearranged. This may be helpful because optimization of the axis order may decrease clutter in the plot. The goal should be to minimize the amount of crossing lines which also improves readability of the plot. The order of axes can also have impact on how user understands the significance of the data because relationships between adjacent dimensions are easier to perceive [28]. Due to the nature of line plots, parallel coordinates plot often suffers from overlapping lines and cluttering that is making the figure unreadable. In order to neutralize this problem, highlighting a specific line with an additional tooltip or label may prove to be a priceless workaround offering the user much better experience when working with the plot. Furthermore, brushing [26] is a form of highlighting where all series based on a specified condition are highlighted 4. In addition, parallel coordinates plot has a 3D variant. In multi-relational 3D parallel coordinates the axes are placed on a circle with a focus axis in the center, allowing a simultaneous analysis between the focus variable and all other variables [29], as shown in figure 5.

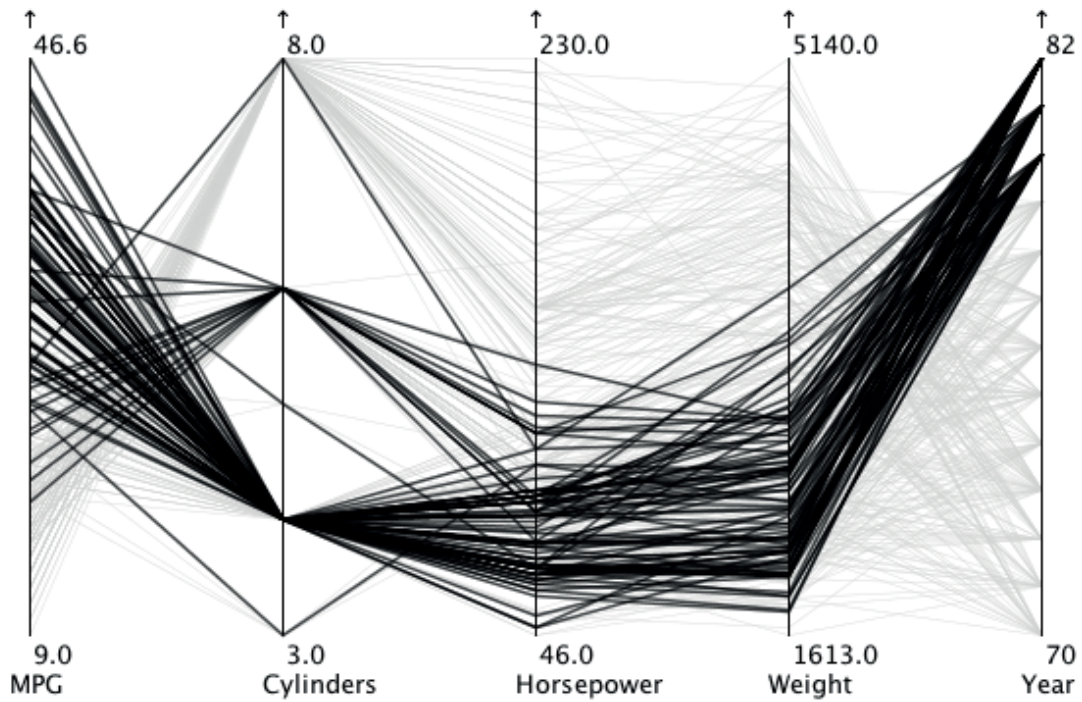


Figure 4: Brushed lines for years 1980 to 1982 on 'year' axis [26].

Disadvantages

Meanwhile, there are several disadvantages for parallel coordinates plot. First, it may display too many objects. This may lead to a large amount of overlapping lines which causes a cluttered and unreadable chart. Usage of axis reordering, highlighting, brushing and other interactive techniques to minimize this issue are highly recommended. Thus, it is hard to make a quick judgement. On the other hand, axis reordering and use of different scales for each axis may complicate interpretations [31]. Further, it may be interesting to perceive lines to be trends or value changes when using the line plots [31]. The purpose of lines is to connect points on the axes and make the relationship between these dimensions more obvious. There are more sources that cause cluttering issues rather than the large amount of lines. One of them is the amount of axes that depends on the width of the plot.

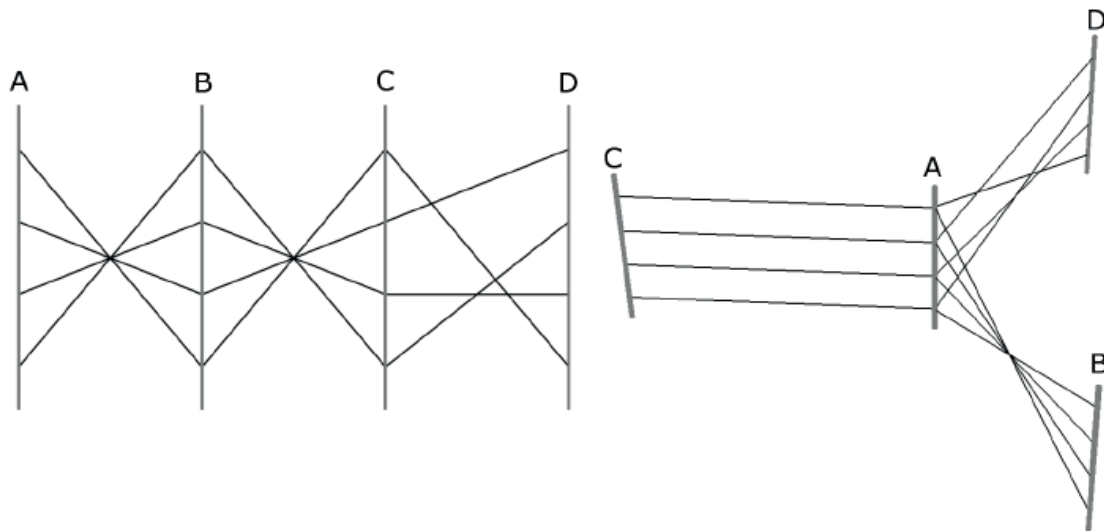


Figure 5: Comparison of parallel plots in 2D variant on the left and 3D variant on the right. [30].

2.5. Radar chart

This visualization method is often seen under many names, such as spider chart, polar chart, star pilots [32]. The aim of the method is to display multiple quantitative variables in a two-dimensional chart. Variables are represented by an array of axes where all start from a center, as shown in figure 6. The arrangement of the axes is radial and maintains equal distance between axes and the length of the axes are also usually the same. The values are placed on respective axis and the resulting set of points can be connected to form a polygon [32]. This polygon represents specific series as individual lines in the parallel coordinates plot with the arguable advantage being that a polygon is more memorable [33].

Advantages

Radar charts are very useful for multivariate analysis and the finding of extreme values. It also provides a quick general overview when certain factors are more prevalent than others. It is possible to display multiple series in the same chart and compare them. It is highly recommended to keep the number of series up to three, otherwise the chart becomes cluttered and will suffer from occlusion. When more series are to be displayed it is advisable to place each series into its own chart and group these charts in multiples, placing them into rows and columns. Then the user obtains the benefit of a compact overview for a larger number of series and the layout is also mobile-friendly [34]. It may also have a different scale for one or more variables. However, the scale should be properly indicated to the user since it will be assumed and expected that all the scales are the same.

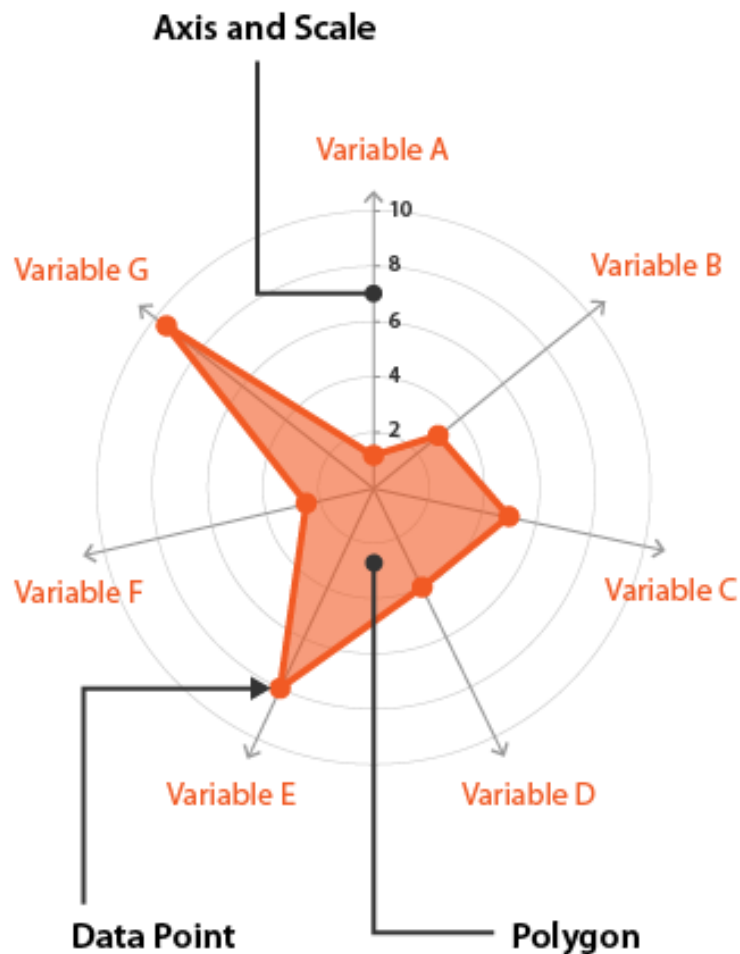


Figure 6: Anatomy of radar chart [32].

Disadvantages

When it comes to radar chart, it is not advised to use the chart as a tool to make trade-off decisions [35]. Similar to the number of displayed series, the large amount of axes will make the chart unreadable and hard to work with [32]. Also, the values on circular layouts are harder to compare. It is much more easier to compare values on bar plot or in other visualizations where values are laid in the same direction of vertical or horizontal axis [34]. Other plots may support order and ranking of variables whereas radar chart does not provide that information clearly [34]. Consider accompanying radar chart with a different visualization entity that supports ranking. When multiple series are displayed the chart will suffer from occlusion. This problem may be reduced by using transparent colors instead of opaque. However, transparent colors introduce tinting, which is hard to read without proper legend. Therefore, variant with no fill color of polygon seems to be the best option at handling occlusion [36].

The visualization does not deal well with data volume. It is not recommended to display more than three series in the same chart. The proportions of the usual shape of the chart makes it possible to stack radar charts horizontally and vertically into grids and comply to the responsive design of modern websites. The variety is restricted to qualitative data. However, the method gains points for decent handling of multi-dimensional data. The velocity handling is poor. The method only provides a snapshot of one moment. In order to observe development over time for new data, a completely new image would have to be composed. Finding patterns and comparing development between two or more completely separate entities may be quite difficult.

2.6. Choropleth

Choropleth is a very popular map-based visualization method widely used to visualize the geographical distribution of data. The main purpose is to display distribution of data and provide an easy way of comparison between geographical areas. In choropleth, the map is divided into regions and areas that are colored, shaded or patterned in relation to a data variable. This relation is usually visible in the form of color progression for each area of the map. This makes it possible to present data values for each geographical area and to observe variation and patterns [37]. In the classed version of choropleth, the variable is classified into several categories where each class has its own color with labels that indicate the range of values it coverage. The recommended amount of used classes is between 3 and 7. There are cases when only two classes are used, namely the political maps which often have only two parties. The more classes we use, the less generalization. However, more classes also means lower readability and higher risk of mistakes [38]. The values are mapped to a gradient consisting of a scale from low to high and each value is represented by a unique color, as shown in figure 7.

Advantages

The choropleth is easy to implement and widely used. It does not take that much time to adjust and use the method to their benefits. Also, it works well with enumeration units such as regions, provinces, or countries, which are suitable for geographical data that rely heavily on these spatial units [38]. Further, choropleth map works well with both numeric and categorical data.

Disadvantages

Choropleth does not offer a proper way to read or compare exact values. Multivariate visualization and display of correlation between values is not supported. Choropleth can be used when observing just one variable. Consider using different methods such as dot plot or scatter plot instead [40]. There is a perception bias towards larger areas. Choropleth maps show us how much geographical area is affected [40]. One has to be careful about making assumptions without enough information and context. For example, choropleths of visualizing voting results may color the map in red and blue depending on which party won in the area. Unfortunately, users are prone to make false assumptions connecting the size of the area to the number of votes.

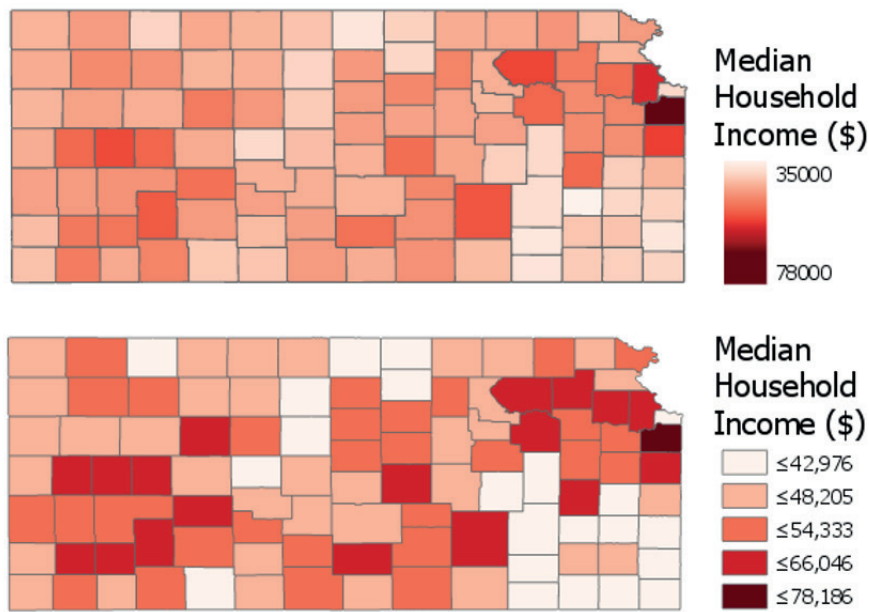


Figure 7: Comparison of color palette for classed and unclassed choropleth [39].

2.7. Streamgraph

Streamgraph is also called stream chart or ThemeRiver [41]. It is a variation of stacked area graphs used to visualize time series data [3]. The main difference is that streamgraph displays values along a central baseline [41]. The graph consists of several individual streams that form layers, an example of streamgraph can be seen in figure 8. These layers are placed within Cartesian axes where the y-axis represents the values of the series and the x-axis indicates time information [42]. The size of each layer is proportionally determined by the value of the respective time series at given time point [41]. Each layer is usually colored to distinguish easily between them and shading may be utilized to describe quantitative characteristics [41].

Advantages

The main advantage of the method is that it achieves two goals that are exclusionary in line-based charts where time series are plotted independently of each other [3, 42]. The first goal lies in the visualization of multiple time series in one chart. The streamgraph is beneficial when used to observe events, patterns and trends that are happening periodically or within specific time frames [41]. The evolution of values is portrayed in an aesthetically attractive way. Layers are usually placed in a manner that the resulting shape resembles a river [41]. This resemblance is useful because it provides an engaging way to illustrate the occurrence of events and trends [3]. Also, the streamgraph provides an efficient way to compare individual layers to the total sum. On the other hand, a layer-to-layer comparison is very difficult [42].

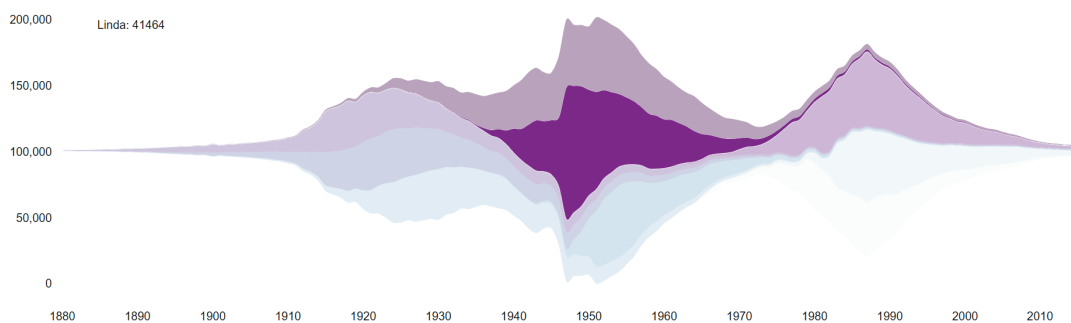


Figure 8: Baby name frequencies in the US from 1880 to 2015 [43].

Disadvantages

The streamgraph is not intended to work with negative values and will not yield proper results in these cases [44]. It is recommended to normalize values and work with compatible range [44]. This issue is caused by the fact that the vertical axis is used to convey size of layers and their stacking [45]. Since is is not applicable to multivariate analysis, the streamgraph works only with a single data dimension [3]. Several factors have an undesirable impact on the aesthetics as described in [42]. Layers may be distorted, placed inadequately or contain too much wiggle. Wiggle is a metric that indicates the degree of visual fluctuation within the graph, as shown in figure 9. The higher it is, the more complex the whole streamgraph is. Optimizations can be made in the baseline selection and layer reordering to minimize these issues. Further, larger amount of series might cause cluttering and readability issues [41]. Use tooltips and highlighting to mitigate this issue. In addition, series with smaller values are prone to be overshadowed by bigger ones and will result in barely noticeable layers. [41]. Use tooltips and highlighting to mitigate this issue.

2.8. Heatmap

The heatmap is a suitable selection when one is looking for a way to examine the correlation of multivariate data [46]. It is often used in a form of a matrix where the value of cells is indicated by their color and shading [46]. Usually, rows portray one category consisting of subcategories and the same applies to columns [46].

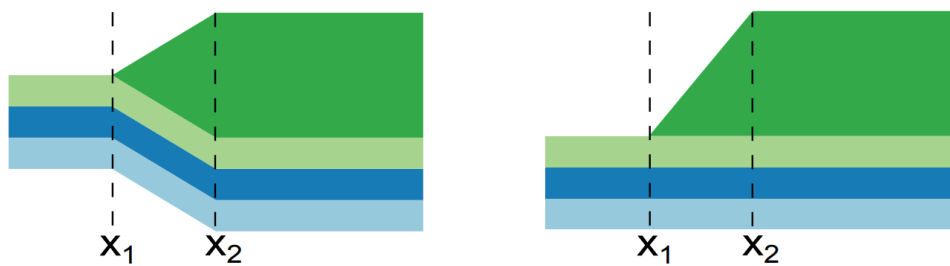


Figure 9: Illustration of different levels of wiggle when selecting baseline with a different algorithm. The left graph contains wiggle in all visible layers whereas the other graph suffers from wiggle only in the green colored layer [42].

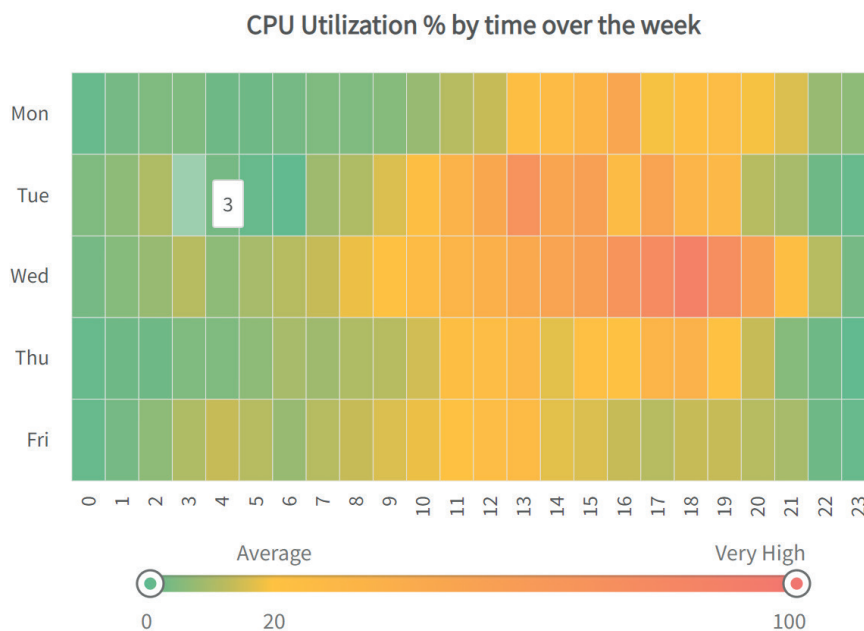


Figure 10: CPU utilization of 23 computers in a network during the working days [47].

Advantages

The heatmap is able to handle both numerical and categorical data [46], as shown in figures 10 and 11. Also, heatmap is useful for showing multivariate data, variance across multiple variables, patterns, general overview and development over time [48]. This method also leverages interactive elements, tooltips and a legend to provide more detailed information. For example, the heatmap in figure 10 provides an option to use the slider in order to highlight cells within a restricted range while other cells go temporarily blank. Furthermore, heatmaps can be used to display the result of hierarchical clustering [49], which is a statistical method that builds a structure which puts similar entities into close proximity. Another alternative is dendrogram. Both heatmap and dendrogram can be placed together in a way that they mirror each other [49].

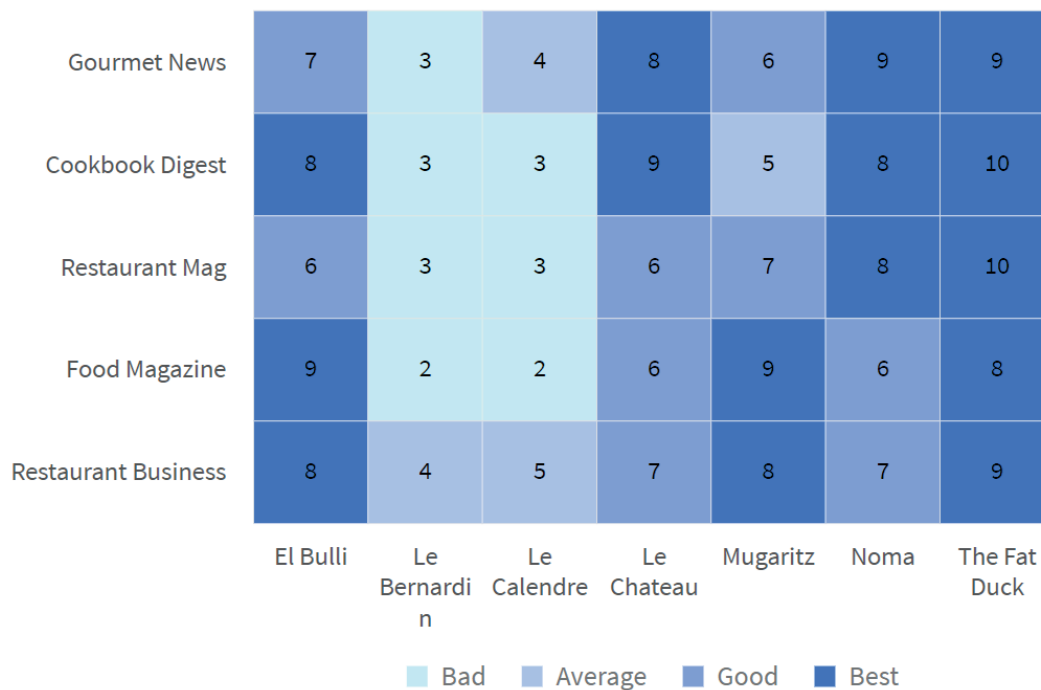


Figure 11: Reviews of restaurant with categorized rating [47].

Disadvantages

It is not advisable to use heatmaps for precision critical analysis. The usefulness of this method lies in its ability to provide a overview. This is caused primarily because of the reliance on color and shading. It can be hard to quickly differentiate certain breakpoints and accurately make conclusions even when given a legend or other supportive tools [46, 50]. Data containing a wide range of values may cause issues with the mapping of values to color scale. It will most likely be necessary to normalize the data [49]. color palette is very important [49]. One needs to be cautious of using more than two distinct hues in the color palette.

2.9. Violin Plot

The violin plot is a visualization method that is frequently used to display the distribution of data. Violin plot combines parts of two similar distribution methods, density plot and box plot (box and whisker plot) [51]. The shape usually consists of a box plot that is enveloped in two density curves (kernel density estimate – KDE) that mirror each other [52]. The anatomy of a box plot can be seen in figure 12.

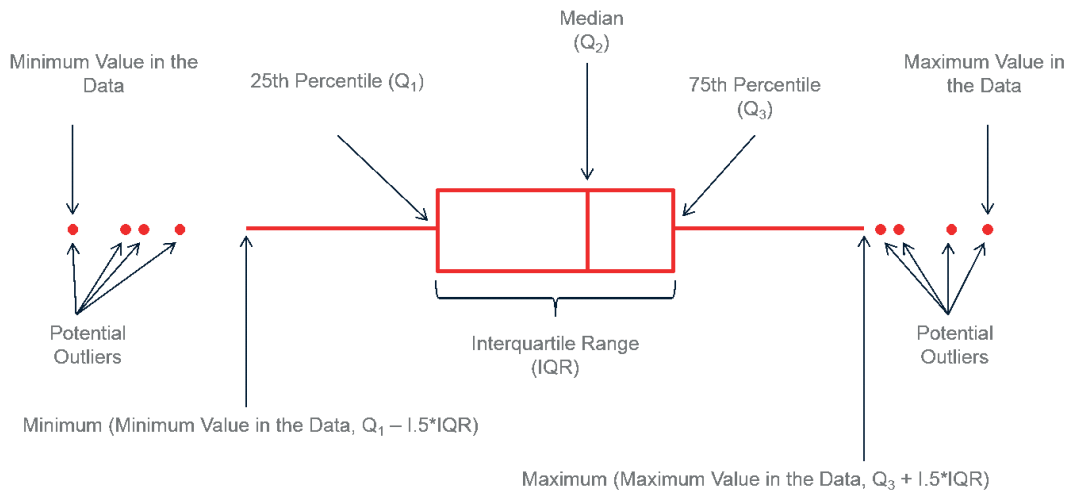


Figure 12: Anatomy of a box plot that may be used in a violin plot [53].

The white dot inside the box plot represents the data median [53]. The thicker line represents the interquartile range [53]. The thin lines extending from top and bottom of the thick line are called whiskers and their purpose is to depict variability outside of the upper and lower quartile [53]. The endings of whiskers represent the upper and lower extreme values (maximum and minimum). Anything beyond is supposed to be considered as an outlier of the dataset [53]. A numerical variable is depicted in several groups where each group has the form of a violin where the width corresponds to the density of data points in the range of the dataset [54].

Advantages

The violin plot provides an efficient way of simultaneously comparing both ranking and distribution of multiple groups [54]. Unlike box plots, it is able to handle multi-modal distributions [51, 55]. It is possible to reorder groups. A frequently used example is to use sorting by median which makes the

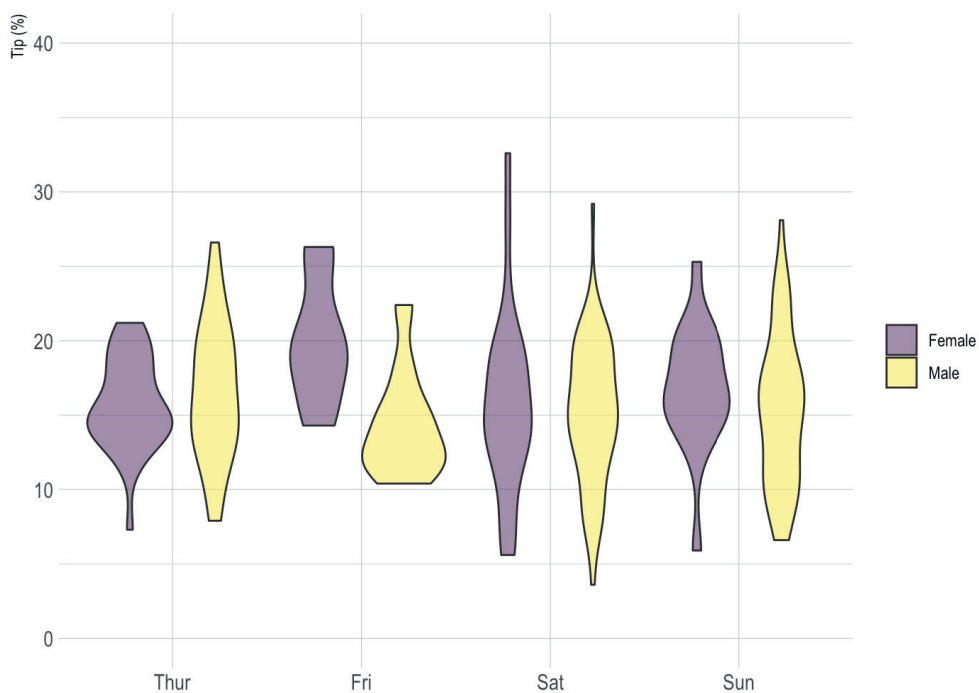


Figure 13: Distribution of restaurant tip during the week with the group distinction between male and female sex [54].

ranking of groups more evident [56]. In addition to grouping, it is also possible to use split violins. The left and right KDE of violin will no longer be mirrored, instead each of them will represent a different group [55, 57]. Likewise, if box plots are used, they should be separated into two individual entities corresponding to the KDE.

Violins can be placed in both vertical or horizontal orientation. This is handy in situations where text labels are too long or when the amount of groups is high and the cumulative length of all labels also causes overlapping issues [56]. Layout development in the area of mobile devices has taught us that when more free space is required, expanding in the vertical axis is the less troublesome way to go [56]. Also, violins are usually displayed with box plots as a part of them. This combination is assumed to be default, however, different objects can be used instead [56]. As shown in figure 14, it is also feasible to overlay the violin with a rug plot, jittered strip plot or swarm plot [56].

Disadvantages

There are also several disadvantages for the violin plot. For example, replacing the box plot as the inner part of a violin can introduce issues with noise and readability [56]. Different entities may provide better insight when the amount of data points is small, but with the increasing size of the dataset it may prove to be distracting and not helpful. If space constraints are an issue or the amount of groups is high, consider using the ridgeline plot [58]. The ridgeline plot saves space and is very useful when the data contains a clear pattern [58, 56]. Similarly, using only the box plot also saves space and it may be a preferable method if statistical summary is of the top priority [56].

2.10. Bullet Graph

The bullet graph was developed and popularised by Stephen Few [59]. The motivation for its introduction was to replace popular meters and gauges used in dashboards that were lacking in certain areas [60]. The bullet graph takes inspiration from traditional bar graph while offering several noticeable upgrades [61]. The bullet graph bar usually consists of five main parts as shown in figure 15. (1) Text label identifies the variable and unit of measurement. (2) Quantitative scale consists of tick marks that convey equal intervals of measure along the bar. (3) Feature measure is a bar indicating reached performance, the primary data. Given its importance, the bar should be visually emphasized. If the quantitative scale starts from value greater than zero, then the featured measure should be displayed in the form of a symbol, preferably a dot or X, instead of a bar. (4) Comparative measure is usually represented in the form of short but very noticeable line that is placed perpendicularly to the feature measure. The purpose of this object is to serve as a target against which the feature measure is compared. The comparative measure should remain visible when surpassed by the feature measure. (5) Qualitative scale is enveloping feature and comparative measures. It is divided into several shaded segments that represents qualitative ranges.

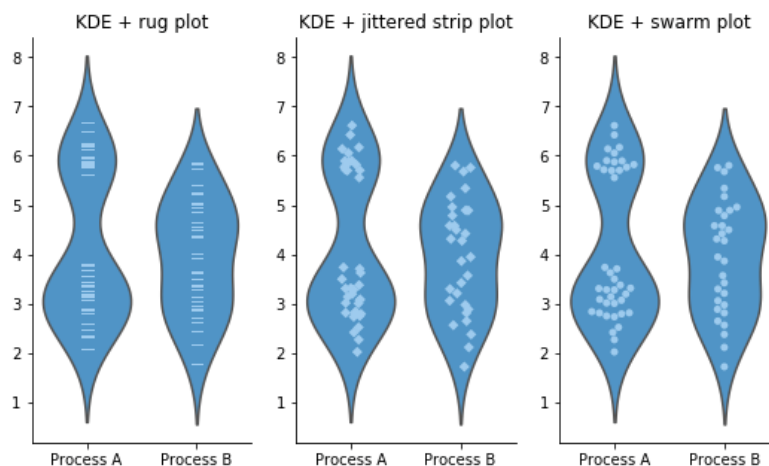


Figure 14: Different layout combinations of violins [56].

Advantages

The bullet graph provides intuitive depiction of performance (gain, loss) against a target value. There are several ways of using color and text labels to convey such information, an example can be seen in figure 16. The main advantage lies in space efficiency which is also the main factor behind utilization in dashboards [60, 62]. Depending on the situation and needs, it is possible to switch between vertical and horizontal orientation of the graph [59, 63]. The shading and hue of qualitative scale can be changed or flipped to indicate whether the state is favorable or highlight specific aspects [64]. Keep in mind though, that qualitative ranges should not be the primary element of the graph. Also this visualization is capable of incorporating future target and predictions [64]. One way to achieve this is to split the bar of feature measure into two distinct color shades. This makes it possible to track both the current progression and the estimated final result.

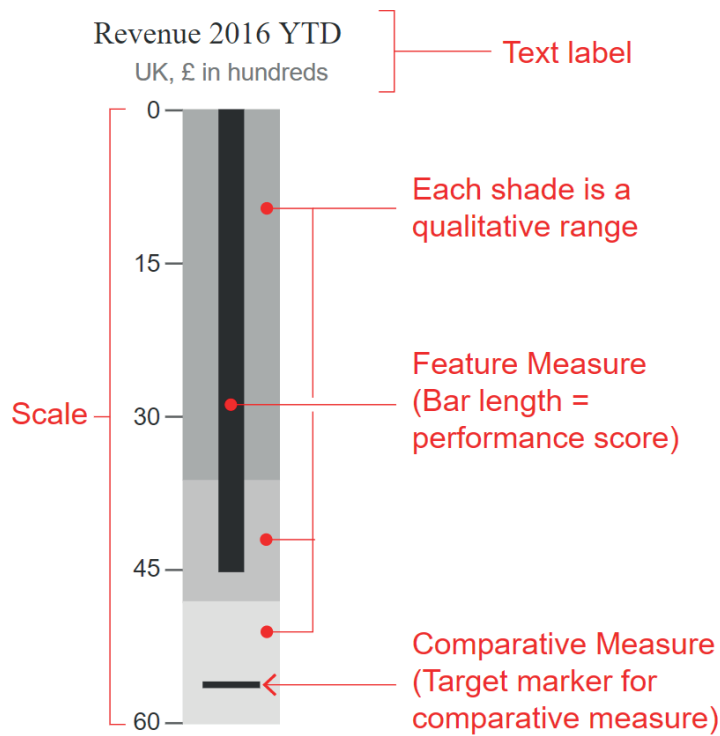


Figure 15: Anatomy of bullet graph [61].

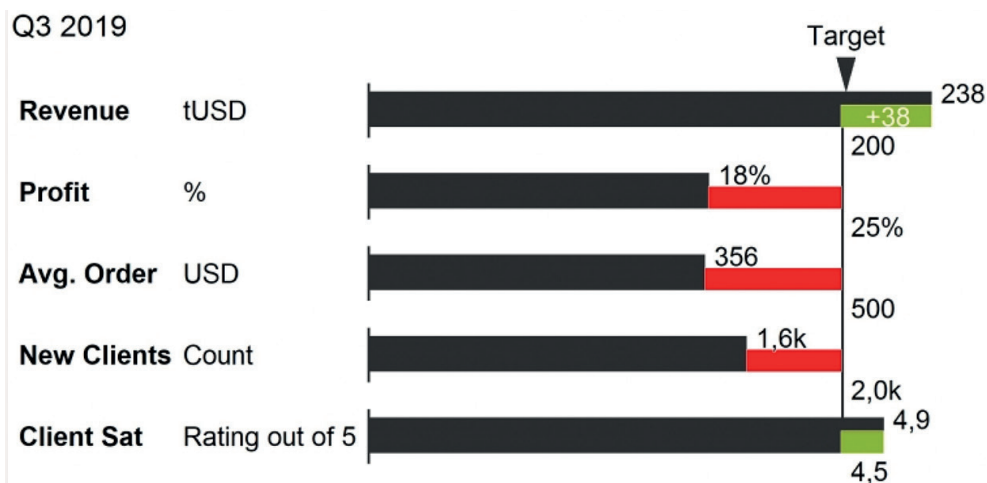


Figure 16: Different visual form of bullet graph displaying highlighted positive or negative variation from target [65].

Disadvantages

In the bullet graph, inappropriate color selection and configuration of qualitative scale can quickly make the graph hard to read and confuse inexperienced users [64]. Also, it is hard to identify the variation of the data value to the target [66]. The method indicates whether the target is reached, however, the variation itself and the amount may not be as quickly recognizable. It is possible to add another bar that would use specific colors to provide a fast indicator of positive or negative development [66], as shown in figure 16.

3. Recommendation Model

Multiple-criteria decision making (MCDM) is a decision tool that allows to analyze complex problems and evaluate the benefits and costs of all alternatives [67]. It also serves as a communication tool and a platform where different requirements can be specified and appointed with preferences and importance. During the selection of appropriate criteria several qualities should be addressed. Some basic qualities are completeness, redundancy and operability [68]. Completeness refers to the importance of including all relevant criteria that may have impact on the specified goal/problem [68]. Redundancy recommends to remove unimportant or duplicate criteria [68]. Operability requires the criterion to be clearly defined in order to make an assessment [68]. Furthermore, it must be possible to judge all alternatives against this criterion [68]. Due to the general nature of choosing the most suitable method, the list of criteria contains quite a large amount of items with a wide range of importance. The large amount of criteria can be structured into big data characteristics (Table 1), data features (Table 2), use scenarios (Table 3) and other characteristics (Table 4).

3.1. TOPSIS

The foundational method used in this model is TOPSIS, which is the abbreviation for Technique for Order of Preference by Similarity to Ideal Solution. The main advantage of TOPSIS is scalability – the procedure retains the same amount of steps regardless of the amount of criteria [69]. Another factor that had impact on the choice was its simplicity. The recommendation model is intended to be used by a web application which will be used by a wide range of users looking for advice. Many MCDM

Big data characteristics

Volume	Amount of records that can be processed and displayed. The criterion is influenced by the ability to highlight or filter desired data and visual elements.
Variety	Amount of attributes (dimensions) that can be visualized and handling
Velocity	Ability to provide interactive experience and handling of incoming and changing data.

Table 1: Classification of criteria in big data characteristics

methods include matrix calculations involving pair-wise criteria comparison [70] that may complicate the process beyond the understanding of a typical user and increase the overall time required to complete

the process over the tolerable limit. Luckily, TOPSIS does not rely on pair-wise comparison [70] and mentioned user unfriendliness can be avoided.

<i>Data features</i>	
Categorical	The method is capable to operate with data of categorical format.
Numerical	The method is capable to operate with data of numerical format.
Mixed num. & cat.	The method operates with a combination of both categorical and numerical data.
Geospatial	The method is capable of processing geospatial data. That may mean geographical coordinates or other data of specific meaning such country names, city names, country codes, etc.
Temporal	The method is capable to operate with time series and data conveying evolution over time. The criterion can be further differentiated by the focus on single or multiple series.
Network	The method is able to handle datasets with notable connections and relationships between individual nodes of data.
Hierarchical	The method is able to handle datasets where data is clustered into aggregated organisation or system.

Table 2: Classification of criteria in data features

TOPSIS was proposed by Huang and Yoon in 1981 [71]. It has been used in a wide range of areas such as supply and management logistics, engineering, marketing management, manufacturing systems and more [69]. TOPSIS begins with the construction of a decision matrix, as shown in equation 1, where each row represents an alternative and each column belongs to a criterion. Alternatives are rated against all criteria and the matrix is filled with initial scores. Two common rating scales are often used in MCDM: relative and ordinal [67]. In relative scale the alternative receives a value conveying the rank order of preference in comparison to other alternatives (e.g. "1 = best", "2 = second best" etc.). In ordinal scale the alternative receives a value based on how a particular interest is satisfied (e.g. a rating from 1 to 5 where "1" is equal to "very bad" and "5" is equal to "excellent").

$$DM = \begin{matrix} & C_1 & C_2 & \dots & C_n \\ \begin{matrix} L_1 \\ L_2 \\ \vdots \\ L_m \end{matrix} & \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{pmatrix} \end{matrix} \quad (1)$$

<i>Use scenarios</i>	
Distribution	The method is capable of displaying frequency of data spread over an interval or group.
Flow/ Sequence	The method is suitable for displaying movement, flow or sequence of data.
Performance/ Ranking	The method is suitable for displaying reached performance, statistical overviews and provides an intuitive way of ranking records across multiple dimensions.
Composition	The method provides detailed insight about individual parts of a bigger overarching whole.
Comparison	The method provides a meaningful and comfortable way to compare individual records or attributes/dimensions of a single record.
Process modeling	The method provides a mechanism suitable for the modeling of a process. These mechanisms are usually in the form of branching/flow, conditions and time constraints.
Relationships	The method can be used to display correlation, connections or other forms of relationship within the data.

Table 3: Classification of criteria in use scenarios

The process is followed by normalization of the decision matrix. This operation will transform various criteria into non-dimensional attributes allowing comparison across all criteria [72], as shown in equation 2.

$$NDM = R_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^m x_{ij}^2}} \tag{2}$$

The next step introduces weights. Usually, criteria have different levels of importance which is the reason why many MCDM methods share this step of prioritizing of criteria. Each rating is multiplied by the attribute weight resulting in weighted normalized decision matrix, as shown in equation 3.

$$V = V_{ij} = W_j \times R_{ij} \tag{3}$$

Afterwards, both positive and negative ideal solutions are determined in equation 4. Beneficial and non-beneficial attributes have to be differentiated as their positive and negative impact is inverted.

$$\begin{aligned}
 &PIS = A^+ = V_1^+, V_2^+, \dots, V_n^+, \\
 \text{where: } &V_j^+ = \{(max(V_{ij}) \text{ if } f \in J); (min(V_{ij}) \text{ if } f \in J')\} \\
 &NIS = A^- = V_1^-, V_2^-, \dots, V_n^-, \\
 \text{where: } &V_j^- = \{(min(V_{ij}) \text{ if } f \in J); (max(V_{ij}) \text{ if } f \in J')\}
 \end{aligned} \tag{4}$$

Visualization method characteristics

Space efficiency	Degree to which the method utilizes space in a responsible manner and its ability to be responsive and scalable across various display devices. There are various interactive techniques that may greatly improve the amount of accessible information on the screen.
Range/variation	The method offers visual information about variation between upper and lower bounds.
Outliers	The method provides an intuitive way to examine data values that are vastly different from usual expectations and patterns, whether it is an individual data point or aggregated group.
Negative values	The method is capable of processing negative values without interfering with positive values. Certain methods can display negative values at the cost of confusion and high risk of misleading the interpretation.
Relative proportions	The method utilizes relative proportions, usually in combination with additional tooltip information, instead of the exact value depiction.

Table 4: Classification of criteria in other characteristics

The procedure continues with the calculation of the separation distance of each alternative to the positive and negative solution, as shown in equation 5.

$$\begin{aligned}
 S^+ &= \sqrt{\sum_{j=1}^n (V_j^+ - V_{ij})^2} & i = 1, \dots, m \\
 S^- &= \sqrt{\sum_{j=1}^n (V_j^- - V_{ij})^2} & i = 1, \dots, m
 \end{aligned}
 \tag{5}$$

Based on the separation distance calculations the relative closeness of each alternative to the ideal solutions is determined as shown in equation 6.

$$C_i = S_i^- / (S_i^+ + S_i^-), \quad 0 \leq C_i \leq 1
 \tag{6}$$

Finally, the relative closeness indicates the ranking of the model, where the higher the value, the better the alternative.

3.2. Rating for decision matrix

In order to perform a ranking calculation based on the TOPSIS method, the initial rating for alternatives and criteria has to be provided in order to construct the decision matrix. The rating is estimated based on the literature review and results are listed in figures 17 and 18. The rating score has a form of abstract value depicting the estimated ability of the visualization to fulfil the criteria. Due to this approach all criteria are considered to be beneficial. There is no need to transform different units into the same

range. Therefore, the rating is given within the range of 0 (poor) to 1 (good) with steps of 0.25 and the requirement of normalization is already satisfied.

3.3. Weights

The next step integrates weights into the decision matrix. Weights are obtained from a combination of basic data analysis of a data sample and survey that user fills in to indicate preference to certain criteria. The weights are translated into a 5 mark range similar to the previously described rating. Some criteria allow only weights of 0 or 1 (true/false).

3.4. Calculation of the ranking

The calculation itself is executed via Scikit-Criteria library [73]. This open source library offers a collection of MCDM methods packed in a scientific Python solution. The library provides an option to use TOPSIS and retrieve variables of the calculation such as the ideal solution, not ideal solution, closeness and the ranking itself. The calculation requires to be supplied with the aforementioned rating matrix, weight list and specification of the criteria impacts.

Method / Criteria	Volume	Variety	Velocity	Time series	One TS	multiple TS	Categoric	Numeric	Mixed Num Cat	Geospatial	Hierarchical	Network	Distribution
bullet graph	1	0.75	0.75	0.5	1	0	0	0	1	0	0	0	0
choropleth	1	0.25	0.75	0	0	0	0	0	1	1	0.1	0	0.5
heatmap	1	1	1	1	0	1	1	1	1	0	0.5	1	0.5
parallel coordinates	0.75	1	0.75	0	0	0	1	0	1	0	0	0	0.25
radar chart	0	0.75	0	0	0	0	1	0	1	0	0	0	0
sankey diagram	1	1	0.5	0	0	0	1	0	1	0	1	1	0.5
streamgraph	1	0	1	1	0	1	0	1	1	0	0.75	0	0.25
treemap	1	0.25	0	0	0	0	1	0	1	0	1	0	0.5
violin plot	1	0.25	0.5	1	1	1	0	1	1	0	0	0	1
sunburst	1	0.25	0.5	0	0	0	1	0	1	0	1	0	0.25

Figure 17: First part of the rating for decision matrix.

Method / Criteria	Flow (Sequence)	Performance (Ranking)	Composition	Space efficient	Range (Variation)	Outliers	Process modeling	Relationships	Comparison	Negative values	Number of variables	Relative proportions
bullet graph	0	1	0	1	1	0.25	0	0	1	1	0.75	0
choropleth	0	0.5	0.25	0.5	0	0.75	0	0	1	1	0.25	0
heatmap	0	0.75	0	1	0	1	0	1	1	0.75	1	0
parallel coordinates	0	1	0	0.25	0	1	0	1	1	1	1	0
radar chart	0	1	0	1	0	1	0	1	0.5	0.25	0.75	0
sankey diagram	1	0	0.5	0	0	0.5	1	0.5	0.5	0	1	1
streamgraph	0	0	0.5	1	0	1	0	0	0.25	0	0	1
treemap	0.25	0.25	1	1	0	0.75	0	0.25	0.75	0	0.25	1
violin plot	0	1	0	0.75	1	1	0	0	0.25	0	0.5	0.5
sunburst	0.5	0.25	1	0	0	0.5	0	0	0.25	0	0.25	0.75

Figure 18: Second part of the rating for decision matrix.

4. System Prototype

We have implemented the algorithm and the system that serves to visualize the recommendation process for the users¹. In the system, the top navigation is comprised of the step description in the form of a ring with an icon inside and a label underneath. Each step can be highlighted with the accent color indicating the overall progress of the form and the process, respectively. The bottom navigation can be comprised of a left-side aligned button with function of going a step backward or a right-side aligned counterpart for going to the next step. The elements of top navigation are clickable and can be used to traverse the multi-step form just as well as the bottom navigation. It is possible to go backwards and forwards, also, the new result can be generated quickly without going through the whole multi-form again. The multi-step form content includes these three steps: dataset upload, criteria and result.

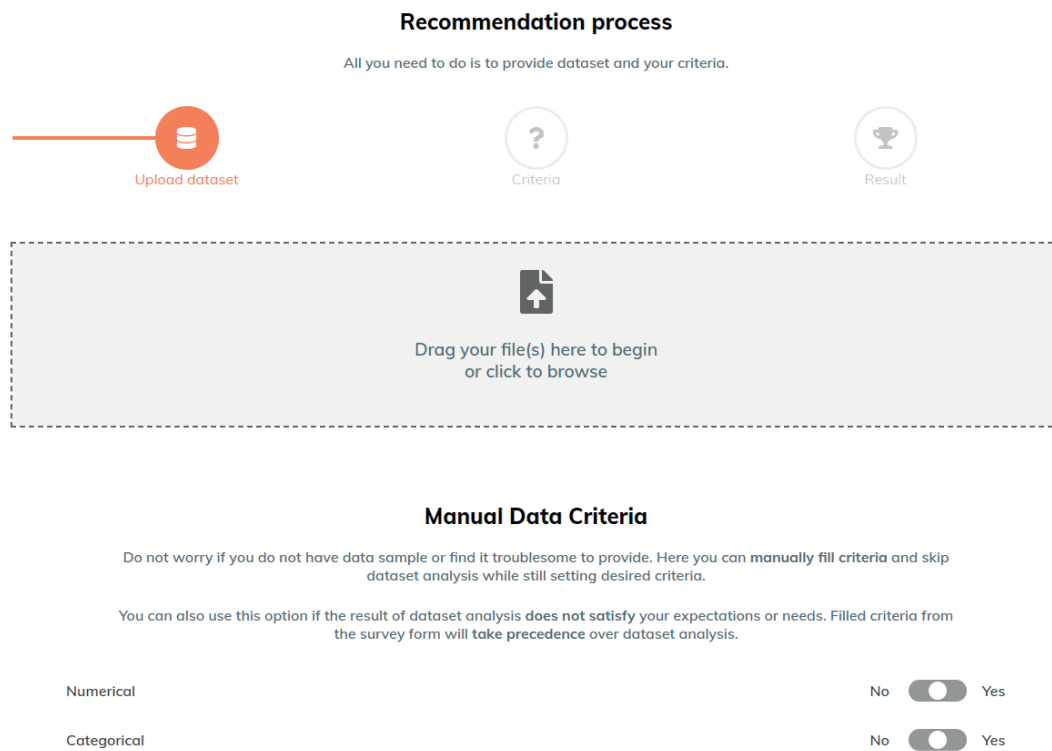


Figure 19: Showcase of the first step in the recommendation process.

The first step, labeled as dataset upload (figure 19), aims to retrieve information about criteria that are closely tied to data properties. These criteria include: numerical, categorical, mixed numerical & categorical, geospatial, temporal, one time series, multiple time series. The first option involves uploading a sample of dataset for analysis. An estimation is conducted and the result is displayed to the user. The second option offers the user a possibility to specify criteria manually. The specification is done via toggle buttons. Both options can be combined, where the manual option enables the user to overwrite result of dataset analysis.

The second step, labeled as criteria, is comprised of several lists of criteria and the user can specify his/her interest in desired criteria. Criteria are either rated in toggle style of true/false options or in a 5 likert scale. The value of 1 signifies zero interest whereas the value of 5 means the criteria is of utmost importance. The last step, labeled as result in figure 20, delivers the result of the recommendation process. The result is presented in a customized three-card layout that is often found on websites with comparison of prices and features of plans, packages or similar paid offerings. The layout depicts the top three choices of the recommendation. Each card starts with an icon of visualization, a text label of

¹ The source code of the system can be downloaded at <https://is.muni.cz/go/n8ngxg>.

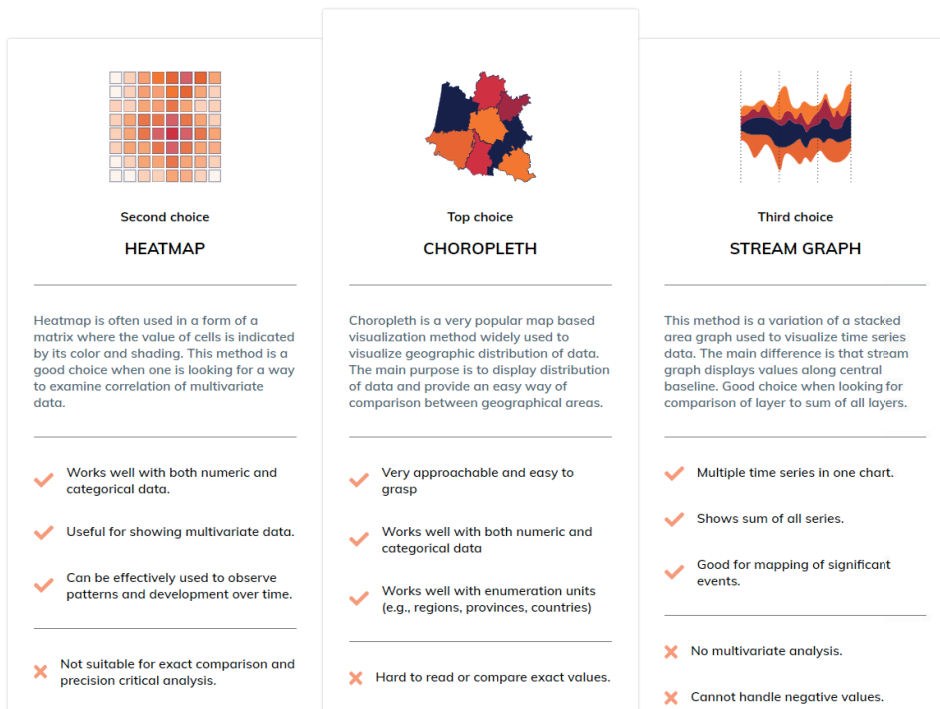


Figure 20: Showcase of the result view in the recommendation process.

the rank order and a visualization name. A short description of the visualization is placed in the middle section. Afterwards, a list of three selected positives and negatives can be found. The last element of each card is a button that links to a detailed description of the visualization.

5. Conclusion

The paper has reviewed and compared different up-to-date visualization techniques for big data. Following the selected visualizations, a set of criteria has been established for the recommendation model that leverages the MCDM approach of TOPSIS to derive suitable visualizations. Based on the proposed model, a system prototype has been implemented. The prototype has shown that the recommendation model can be implemented in practice and help to select data visualization efficiently. The paper has argued that effectively visualizing data is a complex task, while it offers valuable insights for interdisciplinary research. Although a plethora of visualizations have been proposed and each visualization has various advantages, disadvantages and suitable application domains, if the data characteristics and visualization usage are not well understood, the visualization may even complicate the decision-making process instead of helping to obtain insights.

As future works, the recommendation model can be extended to further investigate available possibilities for creation of visualizations in the context of related applications such as Tableau, FusionCharts, Google Analytics, or Sisense. This can be accomplished by enriching each description of visualization. Another future work is to conduct further experimental evaluations by comparing with other recommendation approaches to evidence the effectiveness of the proposed model.

Declaration

This paper is based on the master thesis of “Developing Effective Big Data Analytics by Leveraging Visualization Techniques” from the first author under the supervision of the corresponding author.

References and notes

- [1] M. Ge, V. Dohnal, Quality management in big data, *Informatics* 5 (2) (2018) 19.
- [2] A. Alsaig, V. Alagar, O. Ormandjieva, A critical analysis of the v-model of big data, in: 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), IEEE, 2018, pp. 1809–1813.
- [3] E. Y. Gorodov, V. V. Gubarev, Analytical review of data visualization methods in application to big data, *Journal of Electrical and Computer Engineering* 2013.
- [4] D. Laney, 3d data management, *META group research note* 6 (70) (2001) 1.
- [5] M. F. Uddin, N. Gupta, et al., Seven v's of big data understanding big data to extract value, in: Proceedings of the 2014 zone 1 conference of the American Society for Engineering Education, IEEE, 2014, pp. 1–5.
- [6] L. Wang, G. Wang, C. A. Alexander, Big data and visualization: methods, challenges and technology progress, *Digital Technologies* 1 (1) (2015) 33–38.
- [7] G. Andrienko, N. Andrienko, S. Drucker, J.-D. Fekete, D. Fisher, S. Idreos, T. Kraska, G. Li, K.-L. Ma, J. Mackinlay, A. Oulasvirta, T. Schreck, H. Schumann, M. Stonebraker, D. Auber, N. Bikakis, P. Chrysanthis, M. Sharaf, G. Papastefanatos, Big data visualization and analytics: Future research challenges and emerging applications, in: 3rd Intl. Workshop on Big Data Visual Exploration Analytics (BigVis 2020), 2020.
- [8] M. Lebid, Misleading statistics examples – discover the potential for misuse of statistics & data in the digital age (2018). URL: <https://www.datapine.com/blog/misleading-statistics-and-data>
- [9] R. McCready, 5 ways writers use misleading graphs to manipulate you (2020). URL: <https://venngage.com/blog/misleading-graphs>
- [10] A. Zoss, Data visualization (2019). URL: <https://guides.library.duke.edu/datavis>
- [11] H. R. Nagel, Scientific visualization versus information visualization, in: Workshop on state-of-the-art in scientific and parallel computing, Sweden, 2006, pp. 8–9.
- [12] TABLEAU SOFTWARE, Data visualization beginner's guide. URL: <https://www.tableau.com/learn/articles/data-visualization>
- [13] P. Laubheimer, Treemaps (2019). URL: <https://www.nngroup.com/articles/treemaps>
- [14] K. Robert, Visualization research, part i (2015) URL: <https://eagereyes.org/basics/visualization-research-engineering>
- [15] S. Ribeca, Sunburst diagram. URL: https://datavizcatalogue.com/methods/sunburst_diagram
- [16] J. Stasko, E. Zhang, Focus + context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations, in: IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings, IEEE, 2000, pp. 57–65.
- [17] Y. Holtz, Sunburst (2018). URL: <https://www.data-to-viz.com/graph/sunburst>
- [18] K. Roden, Sequences sunburst (2020) URL: <https://bl.ocks.org/kerryrodden/7090426>
- [19] S. Ribeca, Sankey diagram. URL: https://datavizcatalogue.com/methods/sankey_diagram
- [20] Google Developers, Sankey diagram (2020). URL: <https://developers.google.com/chart/interactive/docs/gallery/sankey>
- [21] ifu hamburg, What is a sankey diagram? (2020). URL: <https://www.ifu.com/en/e-sankey/sankey-diagram>
- [22] M. Schmidt, The sankey diagram in energy and material flow management: Part i: History, *Journal of industrial ecology* 12 (1) (2008) 82–94.
- [23] Phineas, Fifa accounts – my version (2015) URL: <http://www.sankey-diagrams.com/fifa-accounts-my-version>
- [24] Y. Holtz, Sankey diagram (2018). URL: <https://www.data-to-viz.com/graph/sankey>

- [25] FusionCharts, Sankey diagram (2020). URL: <https://www.fusioncharts.com/resources/chart-primers/sankey-diagram>
- [26] R. Kosara, Parallel coordinates (2010). URL: <https://eagereyes.org/techniques/parallel-coordinates>
- [27] Y. Holtz, Parallel coordinates plot (2018). URL: <https://www.data-to-viz.com/graph/parallel>
- [28] S. Rebecca, Parallel coordinates plot. URL: https://datavizcatalogue.com/methods/parallel_coordinates
- [29] J. Johansson, C. Forsell, M. Lind, M. Cooper, Perceiving patterns in parallel coordinates: determining thresholds for identification of relationships, *Information Visualization* 7 (2) (2008) 152–162.
- [30] H. R. Sankey, The thermal efficiency of steam-engines 125 (1896) (1896) 182–212
URL: <https://doi.org/10.1680/imotp.1896.19564>
- [31] Z. Gemignani, Better know a visualization (2010). URL: <https://www.juiceanalytics.com/writing/writing/parallel-coordinates>
- [32] S. Rebecca, Radar chart. URL: https://datavizcatalogue.com/methods/radar_chart
- [33] M. Akmanalp, Mis-employing radar charts to distinguish multidimensional data (2018)
URL: <https://towardsdatascience.com/plotting-in-many-dimensions-382fbd7fe76e>
- [34] Y. Holtz, The radar chart and its caveats (2018). URL: <https://www.data-to-viz.com/caveat/spider>
- [35] FusionCharts, Radar chart (2020). URL: <https://www.fusioncharts.com/resources/chart-primers/radar-chart>
- [36] G. Odds, A critique of radar charts (2011). URL: <https://blog.scottlogic.com/2011/09/23/a-critique-of-radar-charts>
- [37] S. Rebecca, Choropleth map. URL: <https://datavizcatalogue.com/methods/choropleth>
- [38] Axis Maps, Choropleth maps. URL: <https://www.axismaps.com/guide/univariate/choropleth>
- [39] C. Anderson, Making choropleth maps. URL: <https://www.e-education.psu.edu/geog486/node/671>
- [40] L. C. Rost, Choropleth maps (2018). URL: <https://blog.datawrapper.de/choroplethmaps>
- [41] S. Rebecca, Stream graph. URL: https://datavizcatalogue.com/methods/stream_graph
- [42] M. Di Bartolomeo, Y. Hu, There is more to streamgraphs than movies: Better aesthetics via ordering and lassoing, in: *Computer Graphics Forum*, Vol. 35, Wiley Online Library, 2016, pp. 341–350.
- [43] Y. Holtz, Streamgraph (2018). URL: <https://www.data-to-viz.com/graph/streamgraph>
- [44] P. Toffanin, Streamgraph in r [final] (2017). URL: <https://paolotoffanin.wordpress.com/2017/10/03/streamgraph-in-r-final>
- [45] A. Kirk, Stream graph. URL: <http://seeingdata.org/taketime/inside-the-chart-stream-graph>
- [46] S. Rebecca, Heatmap (matrix). URL: <https://datavizcatalogue.com/methods/heatmap>
- [47] FusionCharts, Heat map chart (2020). URL: <https://www.fusioncharts.com/resources/chart-primers/heat-map-chart>
- [48] V. Kesavan, What are heatmaps and what is a good way to use one? URL: <http://daydreamingnumbers.com/blog/what-are-heatmaps-and-what-is-a-good-way-to-use-one>
- [49] Y. Holtz, Heatmap (2018). URL: <https://www.data-to-viz.com/graph/heatmap>
- [50] A. Kirk, Heat map. URL: <http://seeingdata.org/taketime/inside-the-chart-heat-map>
- [51] S. Rebecca, Violin plot. URL: https://datavizcatalogue.com/methods/violin_plot
- [52] L. Ngo, 5 reasons you should use a violin graph (2018). URL: <https://blog.bioturing.com/2018/05/16/5-reasons-you-should-use-a-violin-graph>
- [55] D. Coleman, Box plot with minitab (2015). URL: <https://www.leansigmacorporation.com/box-plot-with-minitab>

- [54] Y. Holtz, Violin plot (2018). URL: <https://www.data-to-viz.com/graph/violin>
- [55] J. Carron, Violin plots 101 (2016). URL: <https://mode.com/blog/violin-plot-examples>
- [56] M. Yi, A complete guide to violin plots (2019). URL: <https://chartio.com/learn/charts/violin-plot-complete-guide>
- [57] H. Pampiglione, Violin plots (2019). URL: <https://ion.icaew.com/data-analytics/b/blogs/posts/violin-plots>
- [58] Y. Holtz, Ridgeline plot (2018). URL: <https://www.data-to-viz.com/graph/ridgeline>
- [59] T. Bodawala, Bullet charts – what is it and how to use it (2019). URL: <https://www.atyantik.com/bullet-charts-what-is-it-and-how-to-use-it>
- [60] FusionCharts, Bullet graph (2020). URL: <https://www.fusioncharts.com/resources/chart-primers/bullet-graph>
- [61] S. Rebecca, Bullet graph. URL: https://datavizcatalogue.com/methods/bullet_graph
- [62] S. Evergreen, Bullet graph (2014). URL: https://www.betterevaluation.org/en/evaluation-options/bullet_chart
- [63] B. Singh, 5 use cases for bullet charts (2017). URL: <http://visualbi.com/blogs/business-intelligence/dashboards/bullet-charts-use-cases>
- [64] S. Few, Bullet graph design specification, Perceptual Edge-White Paper 5.
- [65] Chart-me, C14h bullet graph columns bars. URL: <https://chart-me.com/portfolio-item/c14h-bullet-graph-columns-bars>
- [66] M. Wolff, Bullet graphs. URL: <https://www.chartisan.com/en/bullet-graphs>
- [67] Natural Resources Leadership Institute, Multi-criteria decision analysis. URL: <https://projects.ncsu.edu/nrli/decision-making/MCDA>
- [68] J. S. Dodgson, M. Spackman, A. Pearman, L. D. Phillips, Multi-criteria analysis.
- [69] M. Velasquez, P. T. Hester, An analysis of multi-criteria decision making methods, *International journal of operations research* 10 (2) (2013) 56–66.
- [70] P. Bajec, D. Tuljak-Suban, A framework for detecting the proper multicriteria decision-making method taking into account the characteristics of third-party logistics, the requirements of managers, and the type of input data, in: *Application of Decision Science in Business and Management*, IntechOpen, 2019.
- [71] C.-L. Hwang, K. Yoon, *Multiple attribute decision making—methods and applications* springer-verlag berlin heidelberg, New York.
- [72] P. Mishra, Topsis - a multi-criteria decision making approach (2013). URL: <https://www.slideshare.net/pranavmishra22/topsis-a-multicriteria-decision-making-approach>
- [73] J. B. Cabral, N. A. Luczywo, J. L. Zanazzi, Scikit-criteria: Colección de métodos de análisis multi-criterio integrado al stack científico de Python, in: *XLV Jornadas Argentinas de Informática e Investigación Operativa (45JAIO)- XIV Simposio Argentino de Investigación Operativa (SIO) (Buenos Aires, 2016)*, 2016, pp. 59–66. URL: <http://45jaiio.sadio.org.ar/sites/default/files/Sio-23.pdf>