Data visualization and it's analysis*

Andrejčík Dávid
Slovenská technická univerzita v Bratislave
Fakulta informatiky a informačných technológií
xandrejcik@stuba.sk@

prvý semester 2023

Abstract

Visual analytics is a science based on taking massive loads of data, visualizing parts we need or want to, analysing isolated data and making a conclusion out of it. Considering advanced sensors like, movement tracking or temperature sensors, as well as simulations of many kinds, sky might as well be the limit here. Infrastructure can be improved by tracking common vehicle routs, fire exits are designed based on fire simulations in buildings and even our movements are being analysed from data provided by GPS systems or surveillance cameras. My paper is going to be focused mainly on traffic infrastructure (or other uses if important or interesting) and its improvements using visualization ...

1 Introduction

Nowadays, with urbanization on it's highest, planning and maintaining infrastructure such as highways, traffic networks and public transportation has become more challenging than ever. Thanks to modern sensory systems, GPS and surveillance cameras, improvements can be made. How are data gathered, visualized and then analyzed in general, can be read in section 2. Interesting uses other than traffic related are briefly touched in small section 5. However, main focus is on how traffic network is tracked, data constructed by various methods and afterwards successfully used in field to shorten time spent on road 3 4 ??.

2 Ways of visualization and analysis

2.1 Space time cube

A three-dimensional Cartesian coordinate system is used to create a space-time cube, which is used to represent spatiotemporal.[1] or moving objects. Two of the three dimensions are used as a location plane, which shows an object's position in space as points connected to geographical coordinates on the study area

*Semestrálny projekt v predmete Metódy inžinierskej práce, ak. rok 2023/24, vedenie: PaedDr. Pavol Baťalík

[1] belonging to both space and time (or space-time)

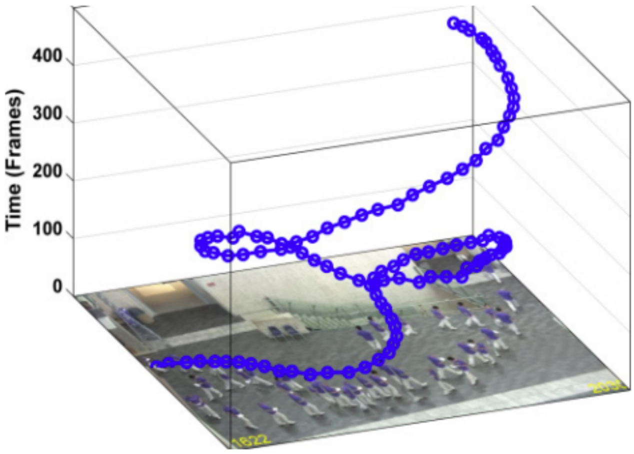


Figure 1: Space time cube. Source [9]

map. The third dimension is the time axis, which shows an object's positions in time as points on the axis. A moving object is represented as a space-time route that connects spatio-temporal locations that the object visits in time, and a spatio-temporal object is represented as a point that associates two spatial coordinates with a temporal coordinate. When only one element is being monitored on a stationary scale or map, for instance, space-time cubes are typically utilised [9] [13].

A brilliant example of a space time cube may be seen in the first image 1. A photograph of a guy who has been caught is stacked on the bottom of axis Z, and blue dots are displayed in a space-time cube where axis X and Y represent the image's position arrays and axis Z represents the number of frames that correspond to the amount of time spent in area vision.

2.2 Graph models

2.2.1 Traj Graph Model

TrajGraph designs a graph, G, which represents a network of roads that cars drive across in a specified amount of time, T. A road section in a city is represented as a vertex in G. A directed edge AB is introduced to road segment G when a taxi goes from road segment A to its related neighbour B. Therefore, G may be utilised to illustrate a network of roads from a vast fleet of taxis navigating city streets.

For the purpose of studying street relevance, streets without taxi service are deemed to be of little significance. Additionally, numerous graphs made for various time frames may be used to illustrate the various traffic data. With this method, users can calculate G periodically every two hours, for example, to analyse the temporal variations of transportation functions over several Gs at once [5].

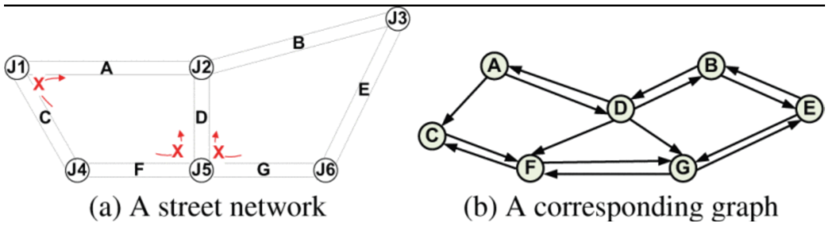
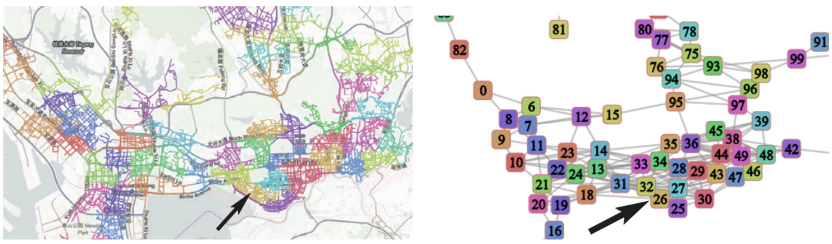


Figure 2: Example of creating TG graph from a street level [5]

2.2.2 Graph partitioning

Graph partitioning is a technique that divides a graph into many segments while meeting predetermined goals and limitations. The most typical restriction is to create partitions with equal chunk sizes, and the most common goal is to reduce the number of edges between the split chunks. A street-level graph of Shenzhen divided into various sections, with colours chosen to represent distinct areas.



Graph partitioning used in Shenzhen. Source: [5]

3 Traffic tracking

3.1 Importance

Modern urbanisation has presented a global problem for transportation management. Taxi are the most active and challenging type of transportation for city officials to manage, especially in medium-to-large cities. When compared to other forms of transportation, such as private vehicles and transportation options like buses and subways, taxis offer a number of distinctive features. First off, taxi cab routes and operation hours are quite flexible and diverse. Second, a number of parties have an interest in their activities, including passengers, taxi drivers, taxi operators, and local government officials.Which is why they are a perfect candidate for monitoring with movement sensors. [8]

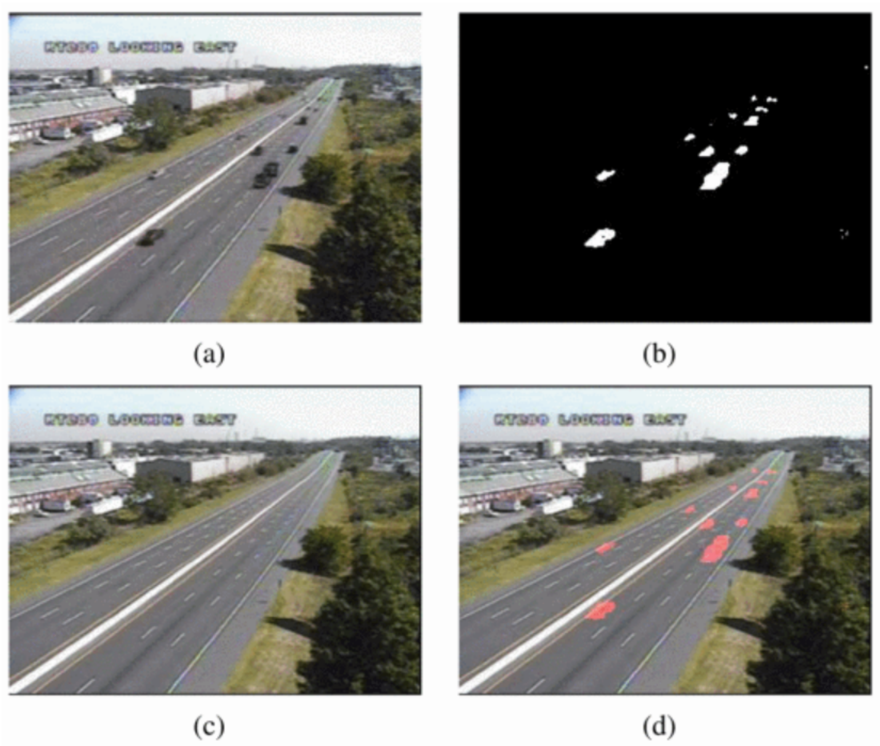
Sensors and measures used:

- Movement sensors - speed cameras, GPS, GSM
- Heat monitoring
- Border controls
- Jam-up reports
- Traffic accidents

3.2 Movement

The widespread availability of Global Positioning System (GPS) and Global System for Mobile Communications (GSM) services, along with the widespread use of smartphones, have made the implementation of tracking systems simple. These days, most smartphones have GPS modules installed, which may be accessed and used by user-friendly mobile application developers. Therefore, using smartphones as location detecting sensors rather than other specialised gear is more practical for vehicle tracking systems. [2]

3.3 Heat-level



Source [11]

Heat-level cameras are one of the ways used to manage oncoming traffic situations. However, not that popular due to their price, these cameras can be efficient in more ways than just traffic density [11].

3.4 Road accidents

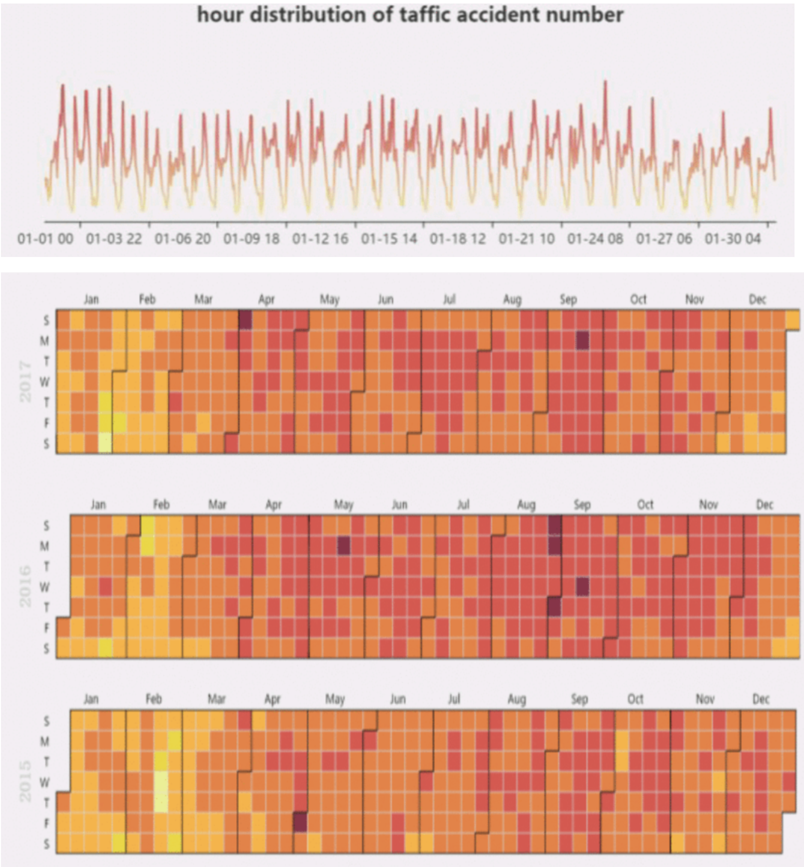
Regretfully, one of the more effective approaches takes traffic accidents into account. The most accurate data are by far those derived from traffic accidents. No road design can eliminate human mistake. Nevertheless, it may be significantly reduced by analysis. Talking about 50 percentage reduction, traffic accidents are ironically preventing more of themselves to happen [7] [6].

Place	YEARLY CRASHES per 100,000 people	USA road ranking (from 1-worst to 50-best)
Mississippi	26.2	11
Arkansas	23.1	8
New Mexico	22.9	5
New York	5.8	45
Hawaii	6.5	42

Figure 3: Comparasion of yearly crashes per x people and road ratings done by NHTSA(higher number means better result). Data in table from [10]

3.4.1 When do they happen

Extracted the traffic accident data in 2017 on one expressway from public security traffic management information system. Visualised charts in the form of a calendar, time matrix, or graphs to more intricately and visually represent the time period during which accidents frequently occur. This data helps mainly to reduce accidents caused by weather and larger traffic movements in some seasons. [3]

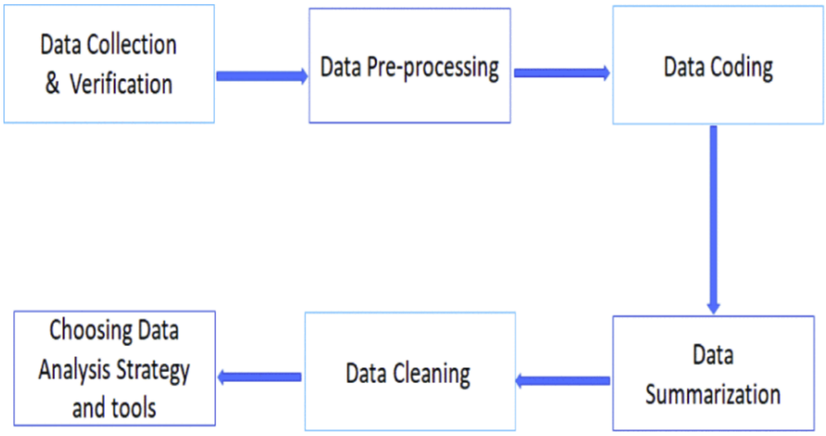


Hour and Date distribution of traffic accidents.(darker-worse) Source: [3]

4 Visualization

Commonly used method:

Using data from road traffic, a prediction model is suggested to examine the accident areas. Data analysis from road accident zones will be done utilising the unsupervised machine learning clustering approach. Maps are used to visualise the data after it has been grouped into various subgroups and analysed based on attributes. For the first phase, the accident data set comes from the public domain. [12]

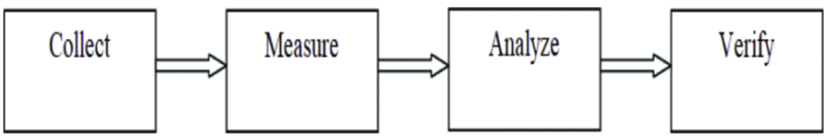


Source of data [12]

Within a certain region, the data collection includes incidents and places close to traffic accidents. Validation of the data set is required prior to pre-processing. To get rid of any potentially redundant or null data, data cleaning should be done. Clean data is used as pre-processing data and supplied as input into the algorithm. The algorithm creates features based on several qualities of the dataset in order to categorise the data. A probability for the processed data is generated by applying the procedure to evaluate the data and based on the features that are obtained. Next, a data visualisation approach is used to the data. The processes for processing the data are shown in diagram 4 above. [12]

4.1 Data Collection and Verification

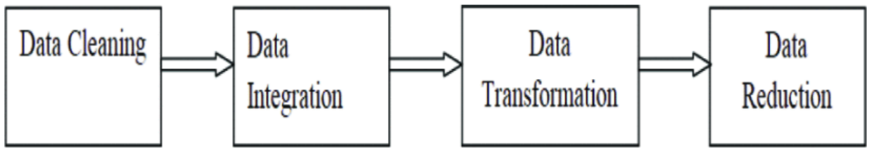
The procedure of obtaining information required by the inquiry and assessing the results is known as the Data Collection technique, and it is shown in diagram 4.1. Verification is the process of examining discrepancies in the information gathered, which is required to gauge advancement. It ascertains if the data was accurately transported from the source to the destination. In order to assess, anticipate the results, and identify patterns, the gathered data is measured, examined, and verified as correct data from the resources. [12]



Source of data [12]

4.2 Data Pre-processing

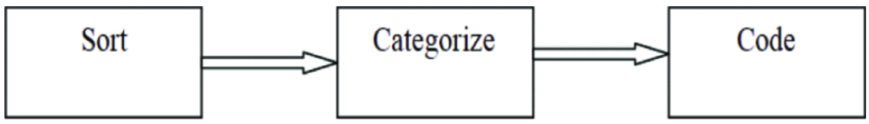
Pre-processing of the data is an essential stage in data analysis. It entails transforming unstructured data into a computer-readable format. We can see this in diagram 4.2. Incomplete raw data may contain mistakes and lack a consistent format. Cleaning out redundant and unneeded data is the first step. The process of merging data from several sources is known as data integration. The purpose of data transformation is to change the shape of data. The last phase involves lowering the amount of data by lowering its storage capacity. [12]



Source of data [12]

4.3 Data Coding

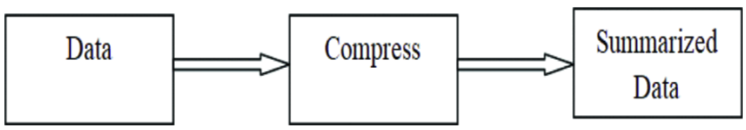
As demonstrated in diagram 4.3, data coding is the process of obtaining codes from observable data. Coding is a method of modifying and improving data. This technique identifies data values and their relationships. Sort and categorise the data in order to investigate more particular concerns. The data is then coded to provide a free-form data structure that can be methodically examined. [12]



Source of data [12]

4.4 Data Summarization

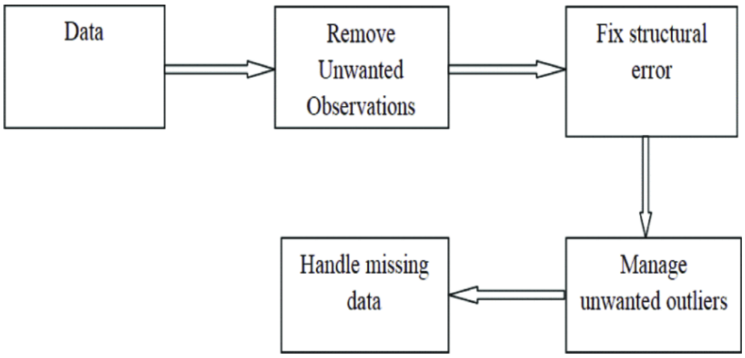
The summarization of created data is known as data summarization. It is the process of reducing data to only the necessary qualities. Diagram 4.4 depicts the data summarization procedure. It is a data value that has been simplified. To obtain valuable information, the entire dataset is summarised. The retrieved data might be displayed in a tabular or graphical fashion. The original data is compressed and called summarised data. [12]



Source of data [12]

4.5 Data Cleaning

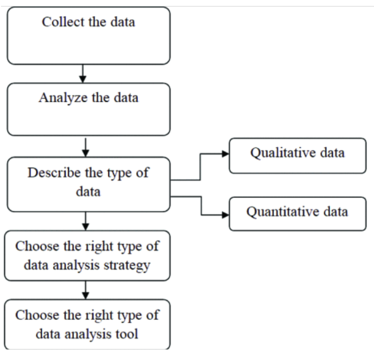
It involves eliminating redundant and noisy data. Combining multiple datasets results in duplicate data, which needs to be cleaned up and should have any superfluous observations removed. Fixing structural issues and eliminating outliers are critical. It's imperative to handle missing data correctly. The data cleaning process is illustrated by the flow diagram 4.5, which addresses error correction, outlier removal, and missing data removal. [12]

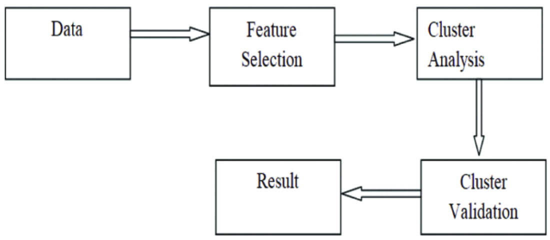


Source of data [12]

4.6 Choosing Data Analysis strategy and tools

The process of gathering and analysing data in order to draw meaningful conclusions is known as data analysis. The two forms of data—qualitative and quantitative—should be taken into account before choosing a data analysis technique. The process of selecting a data analysis strategy and tools is shown in diagram(left) 4.6. When selecting the data analysis tool, the kind of dataset and the intended result must be taken into account. [12]





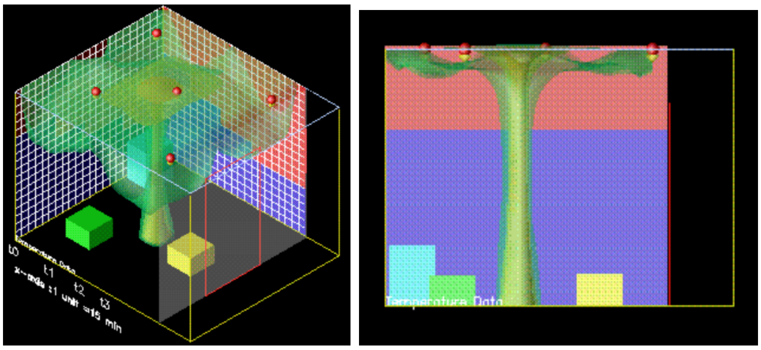
Source of data [12]

Models like cluster, factor, and regression are analysed using a statistical method known as latent class clustering analysis, or LCA for short. This approach entails data collection and analysis. Connections between data are found. Data starts to group together. Latent links between potentially presented observations are shown using Latent Class Cluster Analysis. Group structure is a factor in LCA, which includes discrete latent categorical variables with a binomial distribution. The Latent Class Clustering Analysis (LCA) algorithm's operation is shown in diagram(right) 4.6. [12]

5 Other interesting uses

5.1 Fire exits

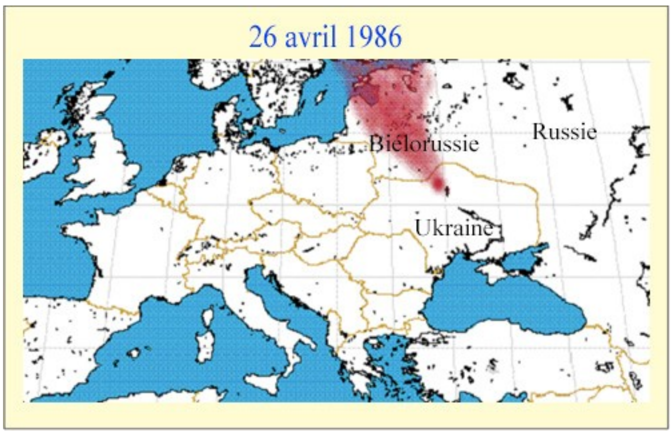
Understanding and forecasting the development and spread of flames has been the main focus of fire management. This has shown itself as a contemporary movement that favours performance above restrictive construction standards.Creating efficient visual representations of fire simulation's component parts for a single room setup. Fire investigators and firefighters are the ones using it. An investigation into linking a database of fictitious fires to the sensors in a real building and producing visual representations of anticipated futures is a long-term objective. [4]

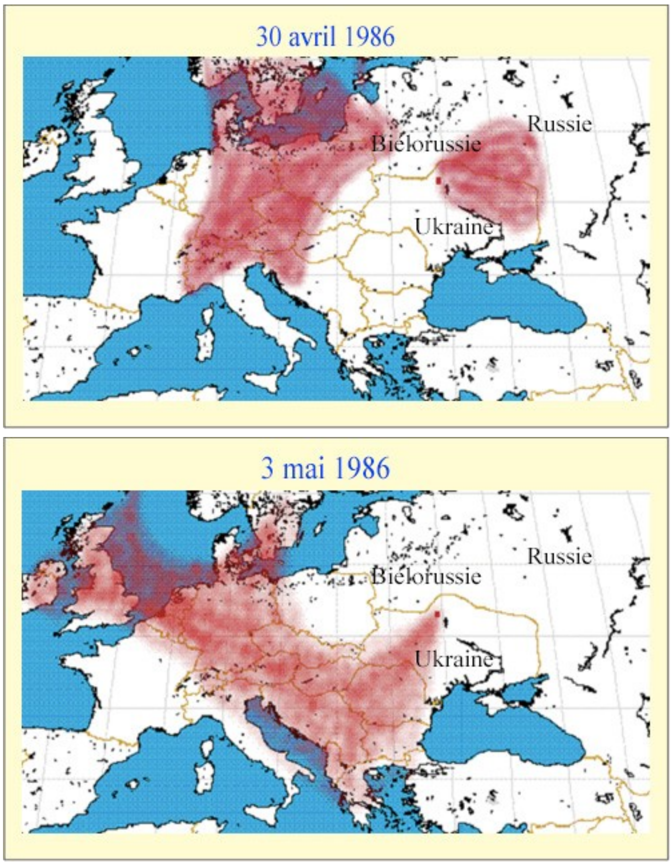


Simulation inside building. Source: [4]

5.2 Radiation spread

Radiation spread can be monitored and therefore analyzed as well as visualized. History examples and safety measures requieres to do so. Altho many factors are in play when considering spread of radiation from say power plant accident, mainly weather and terrain. Taking chernobyl as example: "The Chernobyl cloud has stopped at the french borders"! The broadcast on television and in newspapers of weather maps from April 29 to May 1 suggested that France was escaping the radioactive cloud, due to high pressures in the Azores and favorable winds . . . [14]





Chernobyl's radiation spread. Source: [1]

6 Conclusion

In conclusion, this paper moved into the domain of data visualisation, investigating its potential as a strong tool for presenting complex information and extracting valuable insights. The visualisations described in this work serve not only as representations of raw data, but also as a suggestion, pointing to regions that require the most attention. Furthermore, this paper has illuminated several key findings, providing valuable insights into how to collect, analyse and visualize data. Notably, in traffic, where it shows how road quallity, infrastructure of speed roads and highways and busy season have major inpact on how fast, comfortable and mostly safe the road ahead is.

References

[1] A deep mark in the french collective unconscious. https://radioactivity.eu.com/nuclearenergy/chernobyl_plume [Accessed: (25/11/2023)].

[2] F. N. Ameen, Z. Saeed Mohammed, and A. I. Siddiq. Cost minimization of gps-gsm based vehicle tracking system. In *2018 International Conference on Advanced Science and Engineering (ICOASE)*, pages 256–261, 2018.

[3] A. Fang, X. Peng, J. Zhou, and L. Tang. Research on the map-matching and spatial-temporal visualization of expressway traffic accident information. In *2018 3rd IEEE International Conference on Intelligent Transportation Engineering (ICITE)*, pages 23–27, 2018.

[4] J. Govindarajan, M. Ward, and J. Barnett. Visualizing simulated room fires. In *Proceedings Visualization '99 (Cat. No.99CB37067)*, pages 475–565, 1999.

[5] X. Huang, Y. Zhao, C. Ma, J. Yang, X. Ye, and C. Zhang. Trajgraph: A graph-based visual analytics approach to studying urban network centralities using taxi trajectory data. *IEEE Transactions on Visualization and Computer Graphics*, 22(3):160–169, 2016.

[6] W. Jianjun and Z. Jingjing. Highway traffic accidents and intervening traffic-flow wave model with dynamic space occupancy. In *2009 Second International Conference on Intelligent Computation Technology and Automation*, volume 3, pages 615–618, 2009.

[7] W. Lingtao, Y. Manjuan, T. Chengcheng, P. Daoyue, S. Guohua, Z. Tiejun, and G. Zhanyang. Research on influence extention of two-lane highway intersections based on traffic accident database. In *2010 International Conference on Optoelectronics and Image Processing*, volume 2, pages 244–246, 2010.

[8] S. Liu, J. Pu, Q. Luo, H. Qu, L. M. Ni, and R. Krishnan. Vait: A visual analytics system for metropolitan transportation. *IEEE Transactions on Intelligent Transportation Systems*, 14(4):1586–1596, 2013.

[9] A. H. Meghdadi and P. Irani. Interactive exploration of surveillance video through action shot summarization and trajectory visualization. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2119–2128, 2013.

[10] U. S. D. o. T. NHTSA. National highway traffic safety administration, 2023. `https://www-fars.nhtsa.dot.gov/Main/index.aspx` [Accessed: (18/11/2023)].

[11] H. Shi, H. Ghahremannezhadand, and C. Liu. A statistical modeling method for road recognition in traffic video analytics. In *2020 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, pages 000097–000102, 2020.

[12] S. Subhashini and R. Maruthi. A predictive model for road traffic data analysis and visualization to detect accident zones. In *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*, volume 1, pages 1227–1231, 2023.

[13] T. V. Tran, P. V. Tran, and T. T. Bui. Representing uncertain time on space-time cube for bus movement. In *2013 International Conference on Control, Automation and Information Sciences (ICCAIS)*, pages 353–357, 2013.

[14] D. Williams. Radiation carcinogenesis: lessons from chernoby l- oncogene 27 (suppl 2), s9–s18 (2008), 2009. `https://doi.org/10.1038/onc.2009.349` [Accessed: (25/11/2023)].