A Review on Data Terminology in Visual Analytics Tools

Johanna Schmidt[a] and Milena Vuckovic[b]
VRVis Zentrum für Virtual Reality und Visualisierung Forschungs-GmbH, Vienna, Austria

Keywords: Data Visualization, Data Analysis, Data Engineering.

Abstract: Recent advances in visualization research and related technologies gave rise to several Visual Analytics tools capable of supporting many aspects of a typical data analytics pipeline. More specifically, these tools are showing a promise of a feature-rich environment offering multiple built-in options related to data loading and data management, which are essential initial steps for any data exploratory challenge. In this paper, we review the terms and terminology used to describe data, data parts, and data handling tasks in eighteen commonly used Visual Analytics applications. Throughout this review, we have observed a general lack of standardization of terminology used to describe all related features. Such lack of standardization may affect the overall application potential and increase the complexity when combining different tools, thus creating a user dependency on a specific solution and impeding knowledge exchange.

1 INTRODUCTION

With recent advances in visualization and Visual Analytics (VA) research and related technologies, many promising data analytics frameworks supporting complex inquiries flourished within the past decades (Behrisch et al., 2019). This rise and the increased usage of VA tools prompted the consulting firm *Gartner* to implement a yearly market analysis focusing solely on analytics and business intelligence (BI) products. Gartner's Magic Quadrants (Gartner, Inc., 2023) are a series of market research reports that rely on proprietary qualitative data analysis methods to demonstrate trends, maturity of solutions, and market participants. Software applications like Tableau, Microsoft Power BI, TIBCO Spotfire, and others are listed among these products.

VA applications are considered invaluable assets supporting human analysts in acquiring illuminating insights from their data. However, as increasingly noted by practitioners and visualization researchers (Stoiber et al., 2022), the sometimes non-intuitive user orientation, navigation, and the lack of terminology comprehension of technical terms used to label offered features may deem them impractical and confusing. This may significantly affect the working efficiency reflected in the time required to adapt to and become proficient in using a VA application due to potentially conflicting terminology across frameworks. The lack of standardized across-tools use of terms further leads to a user dependency on a specific application, thus impeding innovation, communication, and knowledge exchange. This is especially relevant considering that many applications do not cover the entire data science workflow, i.e., data discovery, wrangling, profiling, modeling, and reporting (Ruddle et al., 2023). Hence, combining different tools and approaches to achieve the set goals is a typical daily practice for data scientists. One of the essential tasks when using VA applications is *data handling*. Users need to be able to import data into the application, laying out the groundwork for the exploratory data analysis. This relates to tasks such as changing data types, splitting strings, creating custom metrics, assigning relationships between data items based on existing commonalities, and connecting different data sources to establish valuable links to enhance exploratory data analysis.

This paper outlines the intermediate results of an extensive exploratory study on the performance assessment of commercial VA tools. Here, we focus on the practical issues related to the absence of terminology standardization in technical terms used to describe individual features concerning the offered data management, data wrangling, and visualization procedures in VA applications.

[a] https://orcid.org/0000-0002-9638-6344
[b] https://orcid.org/0000-0002-5825-8237

Schmidt, J. and Vuckovic, M.
A Review on Data Terminology in Visual Analytics Tools.
DOI: 10.5220/0012449400003660

In *Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2024) - Volume 1: GRAPP, HUCAPP and IVAPP*, pages 709-716
ISBN: 978-989-758-679-8; ISSN: 2184-4321

2 RELATED WORK

Being viewed as the most work-intensive and cumbersome task, taking up to 80% of the working time (Shrestha et al., 2023), *data wrangling* constitutes one of the first steps in a data science process. Data wrangling can be defined as transforming and mapping data from a raw into another format to make it more fitting and valuable for other purposes, such as analytics or visualization. It comprises, but is not limited to, tasks like file and string parsing, format checks, mitigation of missing and faulty values, data joins, filtering, and possibly sampling.

Milani et al. (Milani et al., 2021) claimed that data workers would greatly benefit from supporting this initial step with visual tools. Other directions in visualization research focused on analyzing data quality (Ruddle et al., 2023) and the use of visualization for data sanity checks (Correll et al., 2019). Data engineering is still an essential part of analytics (Klettke et al., 2021) which also changed drastically over time due to larger and more complex datasets being available today. It is, thus, not surprising that modern commercial VA tools integrated various means for handling and simple data wrangling before visualization and analytics. These features include data import, parsing, data transformations (e.g., date and string formats), and joining different datasets.

Interestingly, in contrast to the importance of the topic, to the best of our knowledge, data handling and/or wrangling have not been a significant part of visualization research yet. As noted by Battle and Scheidegger (Battle and Scheidegger, 2021), users would benefit from interactive solutions to support them in handling data prior to analysis. Emerging from these observations, we aimed to put data handling and wrangling features in VA tools into the focus, to learn from existing approaches, and to identify future research directions.

In contrast to data handling and wrangling, other aspects of the visualization workflow are well-studied. This comprises the visualization design process (Munzner, 2014), the design study stages (Sedlmair et al., 2012), the definition of VA workflows (Gadhave et al., 2022), and the classification of the envisioned modeling methods and analytical tasks to be performed (Andrienko et al., 2018). When focusing on applications and libraries in use, extensive work has been done on evaluating and categorizing visualization capabilities of commercial VA tools (Hameed and Naumann, 2020) and reviewing the usage of visualization techniques in practice (Schmidt, 2022). For the visual representation itself, the grammar of graphics proposed by Wilkinson (Wilkinson, 2005) provides a well-formed foundation for constructing a wide range of graphics. The idea of standardized grammar was picked up by Satyanarayan et al. (Satyanarayan et al., 2017) to propose *Vega-Lite*, a grammar for constructing interactive graphics. A minor degree of this effort has been directed towards the VA tools currently used, their internal organization, and the terminology used to describe all the concerned features. The observed lack of studies in this direction follows similar observations by Zhang et al. (Zhang et al., 2012), who identified the lack of standardization in software components, functionality, and interfaces as a critical issue toward the broad applicability of VA applications.

Within the data science domain, the focus is on the concepts and modeling/mining techniques native to data science workflows. As one example, Kandel et al. (Kandel et al., 2012) provided a formal description of the data science workflow, dividing the process into the five stages of data discovery, wrangling, profiling, modeling, and reporting. In this context, however, the standardization is still evolving, where most of the terminology used is inherited from the fields of statistics and mathematics. Likewise, several emerging associations are mainly working on shaping the profession of data science (i.e., defining roles and titles in data science-related work positions) rather than formulating a set of concrete principles and rules for defining technical terms. Some developments favor the differentiation into data scientists and data engineers (Raj et al., 2019), where data engineers are mainly responsible for maintaining and providing data, including data wrangling. Associations like the *Data Science Association* (Data Science Association, 2023) are currently initiating a more significant movement in this respect by seeking to form a data science standards committee to oversee these developments.

The need for standardization of procedures and terminology is not a new endeavor. This need is reflected in a myriad of contexts such as, for example, language terminology (i.e., tailor-made glossaries), formal concepts and relationships (i.e., ontologies (Booshehri et al., 2021)), (meta)data issues (i.e., FAIR principles (Wilkinson et al., 2019)), clinical practices (i.e., precise nomenclature for diagnoses and treatments), etc. In general, the standardization process creates prevailing norms. It aims to establish a mutual consensus on, among other things, the use of technical terms among a community of experts who represent the field (Gamalielsson and Lundell, 2021). Hence, such a robust system of technical terms plays a vital role in optimizing intellectual and visual communication toward setting up a typical workflow among experts.

3 VISUAL ANALYTICS TERMINOLOGY

We conducted a comparative study between eighteen well-known commercial VA tools (see Table 1). Our selection was informed by our research and experience and by the Gartner review of analytics and business intelligence platforms (Gartner, Inc., 2023). As such, the selected tools constitute a representative sample of the VA tool landscape. The selected VA tools are essentially proprietary solutions where each one offers built-in data connectors, data parser, visualization capabilities, and other features. All tools are well equipped with the features required to support the main activities of a data science workflow - data management, wrangling, and data visualization.

A thorough cross-analysis was conducted across the chosen VA applications, involving numerous user testing sessions that utilized open-source datasets. These sessions have gathered qualitative insights into the user experience, focusing on the clarity and understanding of the terminology encountered during tool interactions, covering all steps of the common data science workflow. We concentrated on the terms and descriptions used to describe data features, tasks, and visualization parts. The study aimed to determine whether common terms are used in the applications or whether terminology differs.

One point we had to agree on was how to structure the process from data import to visualization. In the visualization literature, classification or categorization of the operational steps for obtaining data fitting to be mapped to visual elements is still significantly underrepresented (Walny et al., 2020). In the nested model by Munzner (Munzner, 2014), data engineering tasks are masked by the *Data/task abstraction* step but not outlined in detail. In a data science workflow, the data processing steps are described as *discovery* (finding suitable datasets), *wrangling* (bringing the data into a proper format), and *profiling* (getting to know the data structure).

Walny et al. (Walny et al., 2020) summarize these steps as *data characterization*. The data wrangling process is often split into six parts, similar to Azeroual et al. (Azeroual, 2020), which comprise data exploration, correction, cleaning, validation, and publication. As a conclusion of our literature research and the inconsistencies and lack of a classification of the data engineering process, we summarized our findings and came up with our categorization. It loosely follows the categorization of data wrangling: (i) *data management*, (ii) *data enrichment*, and (iii) *data visualization*. We used this new categorization and studied the terminology used in selected VA applications in the three different stages.

3.1 Data Management

The starting point of our discussion is the stage where users import datasets into VA applications. Importing data into a system requires specific steps to be taken, including opening the file, parsing it, and recognizing the proper data format (numbers vs. strings). Many steps will run automatically, but sometimes user input is required. Manual adjustments and user input may comprise defining the file format (CSV, database, Excel), determining the right deliminator character for CSV files, or defining the date formats in use. All applications provide simple visual representations for viewing at least part of the loaded data for validation.

The overview of used terminologies is given in Table 1. We can observe a variety of expressions and terminologies used along some distinct deviations from the prevalent naming practice that builds on the term ”data” (e.g., data manager, data sources, data files). Some exceptions from this naming convention relate to QlikView where the term ”script file” was employed, which may be due to the equally different approach to handling input data that relies on script execution rather than data in a tabular form. Likewise, Pyramid Analytics used the term ”file sources.” As an interesting detail, we can observe inconsistencies in terminology across solutions released by the same company (i.e., QlikView and Qlik Sense).

Going further, we looked at the common modalities and respective terminologies used for describing original data source data items, data loading features, and evaluation features. Regarding data items, most VA applications adopted the ”field” nomenclature when observing the original imported data. However, we can see other practices in use as well. In some cases, the ”column” term is used, which may be due to the way the input data is represented by the application, which is predominantly in a tabular form. In this tabular form, data attributes are usually represented as columns. In the case of MicroStrategy, a common term for imported and created data items (see also Table 2) was adopted – ”metric or attribute” - which namely reflects the qualitative/quantitative nature of the imported/created data (i.e., non-numeric, numeric). For SAS Visual Analytics, we can further observe a deviation from all identified practices, as the ”data item” term was favored here.

We were mainly interested in usability aspects and the related naming conventions when looking at data loading and evaluation features. Specifically, we looked at data preview options on import and within the VA applications. Surprisingly, only a number of VA applications offer a preview of the data while importing. Where existing, the naming convention of

Table 1: Terminology for data management. This table gives an overview of different terminology used in VA applications when loading data. It shows that the menu item name is not consistent over all applications. Data attributes are in many cases called *fields*, but also *column* and *item* are used. Some applications provide preview of the data, and many added additional information for simple data quality checks.

	Menu item	Input data items	USABILITY		
			Data preview on import (✓, name)	**Data preview in tool**	**Column summary (✓, metrics)**
Domo	data / datasets	column	✓, preview	✓	✓, distribution, quality, statistics
Grafana	data sources	field	-	*	-
HEAVY.AI	data manager	column	-	*	-
IBM Cognos Analytics	data module	column	✓, selected tables	✓	-
Kibana	home / discover	field	✓, data visualizer	*	✓, distribution, statistics
Knowi	data sources	field / metric	-	*	-
Looker Studio	data sources	field	-	-	-
MS Power BI	data	field	✓, navigator	✓	✓, distribution, quality, statistics
MicroStrategy	datasets	metric / attribute	-	✓	-
Pyramid Analytics	file sources	column	-	✓	✓, distribution, statistics
Qlik Sense	prepare/data manager	field	✓, -	✓	✓, distribution, statistics
QlikView	script file	field	✓, file wizard	✓	-
SAS Visual Analytics	data	data item	-	✓	✓, distribution, quality, statistics
Sisense	data	field	✓, add data	✓	✓, distribution, quality, statistics
Tableau	data source	field	-	✓	-
TIBCO Spotfire	files and data	column	✓, import settings	✓	✓, distribution, statistics
Yellowfin	report / data	column / field	✓, preview	✓	✓, distribution, quality, statistics
Zoho Analytics	data sources / data	column	✓, import your data	✓	✓, distribution, quality, statistics**

such preview snippets is quite diverse. However, almost all applications provide a preview once the data is loaded. Hence, if the data types are misclassified (commonly encountered in the case of date-time formats), this can only be adjusted in the applications themselves and not while importing the data. Conversely, only some of the selected applications offer data profiling features with various categories (e.g., distribution, statistics).

3.2 Data Enrichment

Data enrichment can be viewed as an optional step in the data engineering pipeline. In this step, users bring the data into a format that can be used for visualization and analysis afterward. As increasingly reported by data scientists (Azeroual, 2020), data enrichment steps, like adding information or joining datasets, are becoming increasingly important. Joining datasets means that multiple data sources are connected based on existing common fields, and relationships between individual data items are established (similar to relational databases). Table 2 provides an overview of the terms used in the data enrichment stage. We observed a lack of consistency in the general naming practices for the data enrichment stage. Every VA application adopted a custom (proprietary) naming format, which sometimes reflects how the applications process the data (e.g., through scripting). We further looked into the terminology used to describe the established relationships between individual data items (dubbed *data item connections* in Table 2), and a prevalence of the

Table 2: Terminology for data enrichment. This table gives an overview of different terminology used when enriching data with additional information (i.e., creating other data columns). Data enrichment also involves joining multiple datasets. The terms in the table show that the menu items are very inconsistent over all applications, whereas the terms for data connections are pretty consistent. Many applications provide visual scripting interfaces for adding additional data metrics.

			USABILITY		
	Menu item	**Data item connections**	**Visual model**	**Data transformation approach**	**Created data items**
Domo	data / data flow	connection	✓	visual editor	column
Grafana	transform	-	-	scripting	field
HEAVY.AI	SQL editor	-	-	scripting	measure / dimension
IBM Cognos Analytics	data module	relationship	✓	visual editor / scripting	column
Kibana	discover	-	-	scripting	field
Knowi	data transformation	-	-	scripting	field / metric
Looker Studio	*	-	-	scripting	metric / dimension
MS Power BI	model	relationship	✓	visual editor / scripting	column / measure
MicroStrategy	preview	mapping	✓	visual editor / scripting	metric / attribute
Pyramid Analytics	model	relationship	✓	scripting	column
Qlik Sense	prepare / data model viewer	associations	✓	visual editor / scripting	measure / dimension
QlikView	script file / table viewer	associations	✓	scripting	-
SAS Visual Analytics	querry builder**	relationship	✓	visual editor / scripting	measure / category
Sisense	*	relationship	✓	visual editor / scripting	column
Tableau	*	relationship	✓	visual editor / scripting	field / measure
TIBCO Spotfire	data canvas	relationship	✓	visual editor	column
Yellowfin	data transformation	connection	✓	visual editor	field
Zoho Analytics	model	relationship	✓	scripting	column

**Same as in data management. **Provided as a feature of a separate SAS component.*

term ”relationship” is evident. Some of the applications use the term ”associations” or ”connection.” The most considerable discrepancy we observed was in the case of MicroStrategy, where application developers employed the term ”mapping.”

Most of the selected VA use graphical representations such as lines to make it easier for the users to detect and define relationships. The overview of the data relations is often complemented by their cardinality types (i.e., one-to-one, one-to-many, many-to-many, many-to-one) and cross-filter directions (e.g., determines which dataset will be assigned a cross filtering function-–single or both). However, what differs across the applications is the visual clarity of such a resulting data table and the nature of the related relationships. Sometimes a deeper insight into cardinality types and cross filter directions is offered, e.g., depicted through directional arrows (MS Power BI), however more often only the connected data items are visualized (MicroStrategy).

Data enrichment also includes data transformation tasks (e.g., creating custom metrics), which relates to adding additional information to the dataset. In all VA applications this is considered to be a manual task, which requires domain knowledge. Users commonly add custom information through a visual editor enriched by formula expressions or a script editor in the respective applications. Specifically, visual editors relate to user-friendly interfaces supporting data transformation with a visual overview of the origi-

Table 3: Terminology for data visualization. This table gives an overview of different terminology used in the data visualization stage. Data visualization describes the process of creating a visual representation (i.e., chart) of the data. As it can be seen, the menu items used for this stage largely differ among the applications. The used terms (e.g., ”report”, ”dashboard”) reflect the usage of the applications as mostly BI tools. For creating a chart, the terms ”chart”, ”object”, and ”visualization” are mostly used. Others use technical terms like ”widget” or ”view”. Assigning data attributes to charts differs between using axis names and other terms like ”rows” and ”columns”.

			USABILITY		
	Menu item	**Data visualization**	**Line chart**	**Bar chart**	**Scatter plot**
Domo	dashboards	visualization / card	x-axis / y-axis		
Grafana	dashboards	visualization / panel	x-axis / y-axis	axis / value	-
HEAVY.AI	dashboards	chart	dimension / measure	dimension / measure: width	dimension / measure: x-y axis
IBM Cognos Analytics	dashboard	visualization / card	x-axis / y-axis	bars / lenght	x-axis / y-axis
Kibana	dashboard	visualization / panel	x-axis / y-axis		-
Knowi	dashboard	widget	x-axis / y-axis		
Looker Studio	reports	chart	dimension / metric		dimension / x-y metric
MS Power BI	report	visual	axis / values		x-y axis / values
MicroStrategy	dossier	visualization	vertical / horizontal		
Pyramid Analytics	discover	visual	categories / values		x-values / y-values
Qlik Sense	analytics app / sheet	chart	dimension: line / measure: height	dimension: bar / measure: height	dimension: bubble / measure: x-y axis
QlikView	sheet	object	dimensions / expressions		
SAS Visual Analytics	canvas	object	x-axes / y-axes	category / measure	x-axis / y-axei
Sisense	analytics	widget	x-axis / values	categories / values	x-axis / y-axis
Tableau	worksheet / dashboard	view	rows / column		
TIBCO Spotfire	visualization canvas	visualization	column selector: x-y axes		
Yellowfin	dashboard	chart	vertical axis / horizontal axis		
Zoho Analytics	dashboards	chart / view	x-axis / y-axis		

nal dataset (e.g., ”Power Query Editor” in MS Power BI, ”New Custom Column” editor in Sisense). Script editors relate to script-only approaches where visualization of the original dataset is not integrated (e.g., QlikView, HEAVY.AI). The overview of supported features in this regard can be seen in Table 2, dubbed as *data transformation approach*.

Regarding the terminology used to describe created data attribute (i.e., custom metric derived from imported data), the most prominent mixture of terminologies may be observed across selected tools. Frequently, terms in one tool appear to be a mixture of terms used by other tools, as seen in case of Knowi, HEAVY.AI, Looker Studio, and MicroStrategy, where ”measure,” ”dimension,” ”metric”, and ”attribute” seem to be used interchangeably. Also ”category” and ”measure” are used by some tools.

3.3 Data Visualization

In this final step data is transferred to a visual representation. We consider the data visualization stage where all the visuals are coming together to support data exploration in a cohesive and intelligent manner. In our study, we did not go into detail about which types of visualizations the applications offer, as this was already covered by related work. Instead, we focused on the terminology used when creating a graph or a plot in the applications.

Table 3 provides an overview of the terminologies used in the data visualization stage. Similar to a data enrichment space, we can observe various different naming practices for the menu items. While in the majority of cases, the naming reflects the resulting construct of the data visualization stage (e.g., ”report,” ”dashboard”), in others, it seems to reflect the way data visualizations are used (e.g., ”analytics”). Again, we can observe a strikingly diverging practice in MicroStrategy where the term ”dossier” is adopted. Many naming conventions here seem to reflect the usage of the applications as BI tools. Likewise, the terms used to denote a graphical representation of data (i.e., a chart or a graph) vary across the selected solutions. In contrast to what may be understood as a good practice of using a known terminology, such as a ”chart” or a ”visualization,” some solutions departed from this course and embraced a singular concept, as seen in the case of Sisense and Knowi in Table 3 (i.e., ”widget”). The term ”widget” instead describes the technical way how visualizations are usually implemented and integrated into dashboards. QlikView and SAS Visual Analytics employed ”object” for describing a visualization.

Looking at the individual visual representations and the way they handle input data, again, a number of different naming approaches can be observed. While some solutions follow a logical ”x/y axis” data assignment approach (e.g., MS Power BI, Spotfire, Sisense, SAS Visual Analytics), other follow a more geometry-based approach by allocating the data assignment to geometry features of a graphical representation, such as, to a line (in case of line charts), or to a bar and its height (in case of bar charts). This approach is very prominent in case of Qlik Sense and HEAVY.AI. One outlier to both approaches may be observed in case of Tableau, where data assignment follows a tabular approach, having data allocated to either columns or rows, denoting x- and y-axis, respectively. While MS Power BI allows to select data attributes for x- and y-axis, Tableau asks to assign data to ”columns” and ”rows.” As mentioned above, a part of the reason for the observed discrepancies lies in the use of distinct underlying paradigms to manage and represent data (e.g., columns, matrix).

4 RESULTS AND CONCLUSION

The terms used in the stage of data management are still broadly consistent among the tested applications. This consistency is not surprising since users will probably search for the word ”data” in the menu when wanting to import data into the application. Also, the terms used for simple statistics of the loaded data (e.g., ”distribution”) reflect mathematical and statistical measures to check data quality. Significant differences can be observed in the stages of data enrichment and data visualization. Data enrichment is a highly manual stage where users would like to add additional semantic information to the data. Many applications offer graphical representations for defining data relationships and for adding additional details. It is very interesting to see that the way data attributes are mapped to data visualizations is inconsistent in the applications. In many cases, mathematical, statistical, and technical approaches for assigning data attributes are employed. Other tools like Tableau follow a unique approach by defining ”rows” and ”columns.” The different terms for menu items for visualization functionalities reflect how they are used nowadays, namely, as BI tools to create dashboards and reports.

Looking at the implications of our observations, the discussed disparity in existing terminology and data management strategies across the observed VA tools may impose an increased cognitive load on users, impacting both the efficiency and effectiveness of subsequent data analysis. Generally speaking, learning a new set of terms for familiar concepts requires additional time and effort, potentially slowing down the onboarding process and hindering the tool’s effective utilization.

5 FUTURE PERSPECTIVES

In the current phase of the research, direct collaboration with practitioners has not been yet initiated. Nonetheless, it is possible to anticipate certain expectations and perspectives from practitioners centering around the desire to establish a shared terminology within the industry. This would help alleviate potential frustration arising from encountering uncommon jargon that might impede the understanding of VA tools and their functionalities. Overall, substantial implications for user comprehension, efficiency, and the overall analytical performance may be expected.

ACKNOWLEDGEMENTS

VRVis is funded by BMK, BMDW, Styria, SFG, Tyrol, and Vienna Business Agency in the scope of COMET (No. 879730), which is managed by FFG.

REFERENCES

Andrienko, N., Lammarsch, T., Andrienko, G., Fuchs, G., Keim, D. A., Miksch, S., and Rind, A. (2018). Viewing Visual Analytics as Model Building. *Computer Graphics Forum*, 37(6):275–299.

Azeroual, O. (2020). Data Wrangling in Database Systems: Purging of Dirty Data. *Data*, 5(2):50.

Battle, L. and Scheidegger, C. (2021). A Structured Review of Data Management Technology for Interactive Visualization and Analysis. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1128–1138.

Behrisch, M., Streeb, D., Stoffel, F., Seebacher, D., Matejek, B., Weber, S. H., Mittelstädt, S., Pfister, H., and Keim, D. (2019). Commercial Visual Analytics Systems–Advances in the Big Data Analytics Field. *IEEE Transactions on Visualization and Computer Graphics*, 25(10):3011–3031.

Booshehri, M., Emele, L., Flügel, S., Förster, H., Frey, J., Frey, U., Glauer, M., Hastings, J., Hofmann, C., Hoyer-Klick, C., Hülk, L., Kleinau, A., Knosala, K., Kotzur, L., Kuckertz, P., Mossakowski, T., Muschner, C., Neuhaus, F., Pehl, M., Robinius, M., Sehn, V., and Stappel, M. (2021). Introducing the Open Energy Ontology: Enhancing data interpretation and interfacing in energy systems analysis. *Energy and AI*, 5:100074.

Correll, M., Li, M., Kindlmann, G., and Scheidegger, C. (2019). Looks Good To Me: Visualizations As Sanity Checks. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):83–0–839.

Data Science Association (2023). About the Data Science Association. https://www.datascienceassn.org/. [Accessed 2023-05-02].

Gadhave, K., Cutler, Z., and Lex, A. (2022). Reusing Interactive Analysis Workflows. *Computer Graphics Forum*, 41:133–144.

Gamalielsson, J. and Lundell, B. (2021). On Engagement With ICT Standards and Their Implementations in Open Source Software Projects: Experiences and Insights From the Multimedia Field. *International Journal of Standardization Research (IJSR)*, 19:1–28.

Gartner, Inc. (2023). Analytics and Business Intelligence Platforms Reviews and Ratings. https://www.gartner.com/reviews/market/analytics-business-intelligence-platforms. [Accessed 2023-11-20].

Hameed, M. and Naumann, F. (2020). Data Preparation: A Survey of Commercial Tools. *ACM SIGMOD Record*, 49(3):18—-29.

Kandel, S., Paepcke, A., Hellerstein, J. M., and Heer, J. (2012). Enterprise Data Analysis and Visualization: An Interview Study. *IEEE Transactions on Visualization and Computer Graphics*, 18:2917–2926.

Klettke, M., Lutsch, A., and Störl, U. (2021). Measuring Data Changes in Data Engineering and their Impact on Explainability and Algorithm Fairness. *Datenbank Spektrum*, 21:245—-249.

Milani, A. M. P., Loges, L. A., Paulovich, F. V., and Manssour, I. H. (2021). PrAVA: Preprocessing profiling approach for visual analytics. *Information Visualization*, 20(2-3):101–122.

Munzner, T. (2014). *Visualization Analysis and Design*. AK Peters/CRC Press.

Raj, R. K., Parrish, A., Impagliazzo, J., Romanowski, C. J., Aly, S. G., Bennett, C. C., Davis, K. C., McGettrick, A., Pereira, T. S. M., and Sundin, L. (2019). An Empirical Approach to Understanding Data Science and Engineering Education. In *Proceedings of the Working Group Reports on Innovation and Technology in Computer Science Education*, ITiCSE-WGR '19, pages 73—-87, Aberdeen, Scotland, UK. ACM.

Ruddle, R. A., Cheshire, J., and Fernstad, S. J. (2023). Tasks and Visualizations Used for Data Profiling: A Survey and Interview Study. *IEEE Transactions on Visualization and Computer Graphics (Early Access)*.

Satyanarayan, A., Moritz, D., Wongsuphasawat, K., and Heer, J. (2017). Vega-Lite: A Grammar of Interactive Graphics. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):341–350.

Schmidt, J. (2022). Visual data science. In *Data Science, Data Visualization, and Digital Twins*, chapter 6. IntechOpen.

Sedlmair, M., Meyer, M., and Munzner, T. (2012). Design Study Methodology: Reflections from the Trenches and the Stacks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2431–2440.

Shrestha, N., Chopra, B., Henley, A. Z., and Parnin, C. (2023). Detangler: Helping Data Scientists Explore, Understand, and Debug Data Wrangling Pipelines. In *Proceedings of the IEEE Symposium on Visual Languages and Human-Centric Computing*, VL/HCC '23, pages 189–198, Washington, DC, USA. IEEE.

Stoiber, C., Ceneda, D., Wagner, M., Schetinger, V., Gschwandtner, T., Streit, M., Miksch, S., and Aigner, W. (2022). Perspectives of visualization onboarding and guidance in VA. *Visual Informatics*, 6(1):68–83.

Walny, J., Frisson, C., West, M., Kosminsky, D., Knudsen, S., Carpendale, S., and Willett, W. (2020). Data Changes Everything: Challenges and Opportunities in Data Visualization Design Handoff. *IEEE Transactions on Visualization and Computer Graphics*, 26(01):12–22.

Wilkinson, L. (2005). *The Grammar of Graphics*. Springer, San Francisco, CA, USA, 2nd edition.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J., Groth, P., Goble, C., Grethe, J. S., Heringa, J., Hoen, P. A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Jun, Z., and Mons, B. (2019). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3:160018.

Zhang, L., Stoffel, A., Behrisch, M., Mittelstadt, S., Schreck, T., Pompl, R., Weber, S., Last, H., and Keim, D. (2012). Visual analytics for the big data era — A comparative review of state-of-the-art commercial systems. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology*, VAST '12, pages 173–182, Seattle, WA, USA. IEEE.