Literature and Market Analysis

Luciano Melodia, Richard Lenz

Agenda

Project Progress

Project Progress

Project Plan

Year	1. Year				2. Year				3. Year			
Quartal	1	2	3	4	1	2	3	4	1	2	3	4
Process analysis												
Literature analysis												
Market analysis												
Pilot study												
Evaluation												
Conception of the tool												
Prototype implementation												
Documentation												
Milestones	M1		M2/M3	M4		M5	M6		M7			M8

Research Questions for Tool Selection

We investigated tools for *data visualization, processing, cleaning* and *data preparation* for *univariate* and *multivariate* scenarios. Furthermore, *topological tools*, *statistical tools* and tools for *functional dependencies* were investigated.
The following core questions are answered by means of qualitative cataloguing:

RQ1: What are **requirements for exploratory data analysis** tools in industrial use?

RQ2: What **activities** have taken place to exploratory data analysis?

RQ3: What are **popular commercial tools** and what features do they provide?

RQ4: What are **gaps for schema inference** and can we close them?

Legend

Semantics in the following classification

We classify the tools from *different areas of application* in terms of their usefulness for schema inference on gas sensor data. The latter were identified as *time series*. The classification is as follows:

Academic tools and **commercial tools**.

- Important functionalities.
- Suitable tools.

✓Feature support.

(✓) Feature support in development.

Academic Tools

Name	Year	Distinguish attributes	Univariate analysis	Bivariate analysis	Multivariate analysis	Missing values	Outliers	Feature engineering	Scalability	Interpretability	Reduced expertise	User friendliness
DataScope [19]	2018	✓	✓	✓			✓	✓				
DataSite [8]	2019		✓	✓					✓	✓		
Duet [25]	2018	✓	✓	✓		✓				✓	✓	
FastMatch [29]	2018		✓						✓			
InfoNice [52]	2018		✓	✓				✓				✓
Keshif [56]	2018	✓	✓	✓	✓			✓	✓			
Northstar [24]	2018		✓	✓						✓	✓	
Podium [51]	2018	✓	✓	✓	✓			✓	✓			✓
RCLens [27]	2018	✓	✓	✓	✓			✓	✓		✓	

Name	**Year**	Distinguish attributes	Univariate analysis	Bivariate analysis	Multivariate analysis	Missing values	Outliers	Feature engineering	Scalability	Interpretability	Reduced expertise	User friendliness
Taco [35]	2018		✓					✓				
VisComposer [32]	2018	✓	✓	✓				✓			✓	
Voder [47]	2018		✓	✓			✓	✓	✓		✓	✓
Zenvisage [45]	2016	✓	✓	✓			✓				✓	✓
Analyza [10]	2017	✓	✓	✓	✓			✓	✓		✓	✓
ChartAccent [41]	2017		✓	✓				✓		✓		
GaussianCubes [53]	2017	✓	✓	✓	✓	✓		✓	✓			
HindSight [12]	2017		✓	✓				✓	✓			✓
MyBrush [23]	2018		✓	✓		✓	✓	✓				

Name	Year	Distinguish attributes	Univariate analysis	Bivariate analysis	Multivariate analysis	Missing values	Outliers	Feature engineering	Scalability	Interpretability	Reduced expertise	User friendliness
VisFlow [57]	2017		✓	✓	✓		✓	✓	✓	✓		✓
Voyager 2 [54]	2017	✓	✓	✓	✓		✓	✓			✓	✓
AggreSet [54]	2016		✓	✓					✓			✓
DimScanner [55]	2016	✓	✓	✓	✓			✓	✓			
ForeCache [3]	2016		✓	✓	✓			✓	✓		✓	✓
VisTrees [11]	2016		✓	✓				✓	✓			
SeeDB [50]	2015	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓
Sketch [6]	2015		✓	✓				✓	✓			
Bertifier [38]	2015	✓	✓	✓			✓		✓			

Name	Year	Distinguish attributes	Univariate analysis	Bivariate analysis	Multivariate analysis	Missing values	Outliers	Feature engineering	Scalability	Interpretability	Reduced expertise	User friendliness
Domino [15]	2014	✓	✓	✓	✓	✓		✓	✓			
Ellipsis [43]	2014		✓	✓			✓	✓				
Lyra [44]	2014	✓	✓	✓			✓					
PanoramicData [58]	2014	✓	✓	✓			✓					✓
Prog-Insights [48]	2014		✓	✓	✓			✓	✓			
UpSet [26]	2014		✓	✓					✓			
ExPlates [20]	2013	✓	✓	✓	✓			✓				✓
imMens [28]	2013		✓	✓	✓		✓		✓			
LineUp [16]	2013		✓	✓	✓	✓	✓	✓	✓			✓

Commercial Tools

Name	Year	Distinguish attributes	Univariate analysis	Bivariate analysis	Multivariate analysis	Missing values	Outliers	Feature engineering	Scalability	Interpretability	Reduced expertise	User friendliness
Alteryx [42]	2014	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓
Tableau [34]	2014	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓
Domo [31]	2014	✓	✓	✓	✓						✓	✓
Watson Analytics [21]	2015	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MS Power BI [39]	2014	✓	✓	✓	✓	✓	✓	✓	✓		✓	
QlikView [13]	2014	✓	✓	✓				✓			✓	
Sisence [46]	2014	✓	✓	✓	✓			✓	✓		✓	✓
Tamr [17]	2013	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓
Talend [5]	2012	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓

Topological Summaries

Research Tools

Name	Year	Programming language	Univariate analysis	Bivariate analysis	Multivariate analysis	Statistics	Representations	Persistent homology	Scalability	Manifold reconstruction	Reduced expertise	Topological descriptors
Gudhi [30]	2014	C++	✓	✓	✓	✓	✓	✓	✓	✓		✓
Dionysus [33]	2007	C++	✓	✓	✓	✓	✓	✓	✓	✓		✓
Phat [4]	2017	C++	✓	✓	✓	✓	✓	✓	✓	✓		✓
Aleph [2]	2016	C++	✓	✓	✓	✓	✓	✓		✓		
TopToolKit [49]	2017	Python	✓	✓	✓	✓	✓	✓		✓	✓	✓
Kohonen [22]	2014	Python	✓	✓	✓		✓			✓		
JavaPlex [1]	2014	Java	✓	✓	✓	✓		✓	✓	✓		
TDAmapper [37]	2015	R	✓	✓	✓	✓			✓	✓		

Functional Dependencies

Research Tools

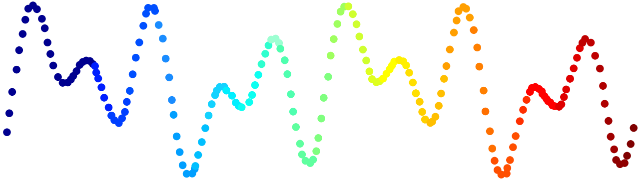
Name	Year	Column statistics	Simple column statistics	Column similarity	CFDs	CINDs	Denial constraint discovery	Candidate key discovery	ETL	Machine learning	Rule based	Data quality
Bellman [9]	2014	✓		✓								✓
Potters Wheel [40]	2014	✓	✓						✓			✓
Data Auditor [14]	2014				✓	✓					✓	
RuleMiner [7]	2014						✓				✓	
MADLib[18]	2014	✓	✓							✓		
Metanome [36]	2015				✓	✓		✓	✓		(✓)	✓

Commercial Tools

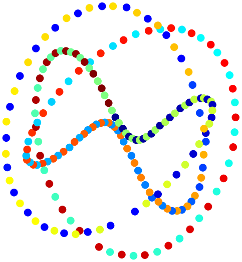
Name	Statistics	Patterns	Uniques	CFDs	MC dependencies	Text profiling	Histograms	ETL	Machine learning	Rule based	Data quality
DQ Analyzer	✓	✓	✓								
InfoSphere	✓	✓			✓						
Data Quality	✓	✓	✓	✓		✓	✓				✓
SQL Server DP	✓	✓	✓	✓	✓			✓			
Enterprise Data Quality	✓	✓	✓	✓	✓				✓		✓
Adaptive Data Preparation	✓						✓				✓
Information Steward	✓	✓	✓	✓	✓			✓			✓
Enterprise / Hunk	✓	✓					✓	✓			
Data Profiler	✓	✓	✓					✓	✓		✓
Trifacta	✓	✓									✓

Illustrating the Problem

Transformation from $f(t)$ into $SW_{M,\tau}f(t)$



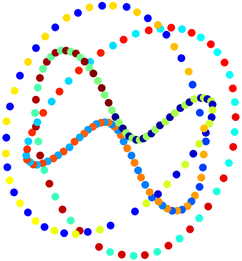
$\longrightarrow$



A 1-dimensional function $f(t)$ embeds into $\mathbb{R}^2$.

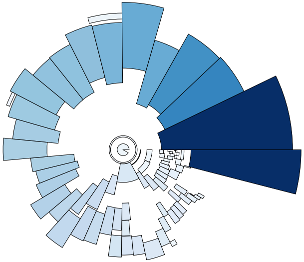
A 1-dimensional planar function $SW_{M,\tau}f(t)$ embeds into $\mathbb{R}^3$.

Computing Persistent Homology



A 1-dimensional function $f(t)$ embeds into $\mathbb{R}^3$.

$\longrightarrow$



Persistence ring developed by Bastian Rieck.

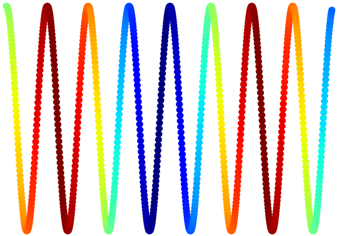
Set up

We choose a metric space $(\mathbb{M}, d)$.
We have given a time series $f : S \subset \mathbb{N} \to \mathbb{M}$.

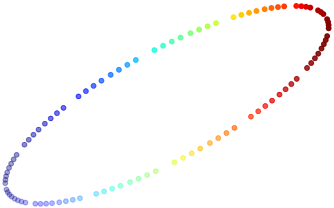
Example:

$F : [a, b] \subset \mathbb{R} \to \mathbb{R}$, is evaluated at
$a \leq t_1 < t_2 < \cdots < t_N \leq b$.

Embedding Periodic Functions



$\longrightarrow$



1-dimensional periodic function $f(t) \hookrightarrow \mathbb{R}^2$

$f(t) = cos(5t)$.

A curve that can be identified with S^1 embedded into $\mathbb{R}^3$.

Embedding Periodic Functions

Given are

$\tau \rightarrow$ step-size or delay,
$M\tau \rightarrow$ window size,
$M+1 \rightarrow$ embedding dimension.

$$\mathrm{SW}_{M,\tau} f(t) = \begin{bmatrix} f(t) \\ f(t+\tau) \\ \vdots \\ f(t+M\tau) \end{bmatrix} \tag{1}$$

Takens' Embedding Theorem

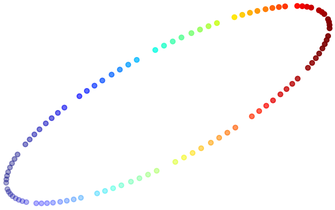
Let $\mathbb{M}$ be a smooth *n*-dimensional Riemannian manifold.
It is a generic property of $\varphi \in \mathrm{Diff}^{\infty}(\mathbb{M})$ and $f \in C^{\infty}(\mathbb{M}, \mathbb{R})$ that

$$\mathbb{M} \rightarrow \mathbb{R}^{2n+1}, \tag{2}$$

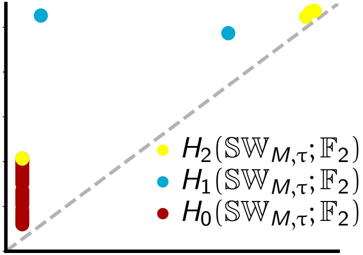
$$x \mapsto \left(f(t), f \circ \varphi(t), \cdots, f \circ \varphi^{2m}(t)\right) \tag{3}$$

is an embedding.
Taking the Fourier transform of a possibly periodic underlying function of a point set yields thus a smooth embedding.
Fourier series can be identified with curves on an *n*-torus.

Computing Persistent Homology



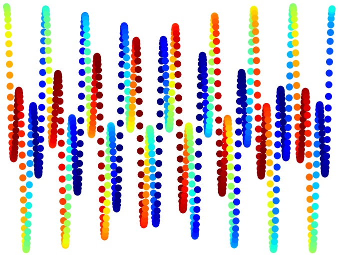
$\longrightarrow$



1-dimensional manifold embeds into $\mathbb{R}^3$

Persistence diagram tracks representatives of homology groups along filtrations.

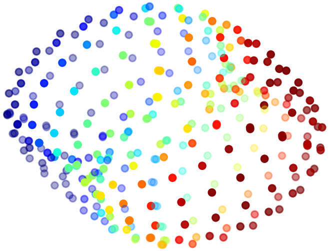
Embedding Quasi-periodic Functions



1-dimensional quasi-periodic function

$f(t) \hookrightarrow \mathbb{R}^2$, $f(t) = cos(t) + cos(\pi t)$.

$\longrightarrow$



Curve dense on a torus in $\mathbb{R}^3$.

Embedding Quasi-periodic Functions

Given a function $f : \mathbb{R} \to \mathbb{C}$ of the form $f(t) = \sum_{n=0}^{N} c_n e^{i\omega_n t}$.
The numbers $c_0, \cdots, c_N \in \mathbb{C} \setminus \{0\}$.
The numbers $\omega_0, \cdots, \omega_N \in \mathbb{R}^+$.

$1, \omega_0, \cdots, \omega_N$ are linearly independent over $\mathbb{Q}$.
The numbers are called incommensurate over the integers, if they can't be expressed as a ratio of integers, such as irrational and transcendental numbers.

Embedding Quasi-periodic Functions

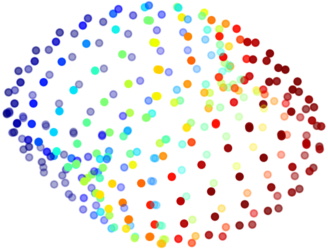
For $M \in \mathbb{N}$ and $\tau > 0$ we have

$$\mathrm{SW}_{M,\tau} f(t) = \begin{bmatrix} f(t) \\ f(t+\tau) \\ \vdots \\ f(t+M\tau) \end{bmatrix} = \sum_{n=0}^{N} c_n e^{i\omega_n t} \cdot \begin{bmatrix} 1 \\ e^{i\omega_n \tau} \\ \vdots \\ e^{i\omega_n M\tau} \end{bmatrix} \tag{4}$$

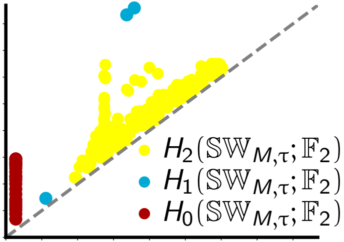
$$= \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{i\omega_0\tau} & e^{i\omega_1\tau} & \cdots & e^{i\omega_N\tau} \\ \vdots & \vdots & \ddots & \vdots \\ e^{i\omega_0 M\tau} & e^{i\omega_1 M\tau} & \cdots & e^{i\omega_N M\tau} \end{bmatrix} \cdot \begin{bmatrix} c_0 e^{i\omega_0 t} \\ c_1 e^{i\omega_0 t} \\ \vdots \\ c_N e^{i\omega_N t} \end{bmatrix} \tag{5}$$

$$= \Omega_f \cdot x_f(t). \tag{6}$$

Computing Persistent Homology



$\longrightarrow$



1-dimensionalal curve on a torus embeds into $\mathbb{R}^3$.

Persistence diagram tracks representatives of homology groups along filtrations.

Kronecker's Theorem

For $c \in \mathbb{C}$ let $S_c^1 = \{z \in \mathbb{C} \mid |z| = |c|\}$ and let

$$x_f(t) = \begin{bmatrix} c_0 e^{i\omega_0 t} \\ c_1 e^{i\omega_0 t} \\ \vdots \\ c_N e^{i\omega_N t} \end{bmatrix}. \tag{7}$$

If $1, \omega_0, \cdots, \omega_N$ are linearly independent over $\mathbb{Q}$, then $\{x_f(t) \mid t \in \mathbb{Z}\}$ is dense on $\mathbb{T}^{N+1} = S_{c_0}^1 \times \cdots \times S_{c_N}^1$.

If $0 \leq \tau \cdot \max(\{\omega_i\}_{1,\cdots,n}) < 2\pi$ then Ω_f is of full-rank. Moreover, if in addition $M \geq N$ then $\mathbb{SW}_{M,\tau} f = \mathsf{SW}_{M,\tau} f(\mathbb{Z})$ is dense in an $(N+1)$-torus.

Inference Problem

1. Each signal corresponds to exactly one sensor.
   This forms an entity $\bullet$-sensor and a set of all such signals $\mathcal{S}^{\bullet}$.
2. Each component of a turbine is an entity, represented by a subset of $\mathcal{S}^{\bullet}$.
3. Each turbine is part of a powerplant consisting of multiple components.
4. This is a nested structure of sets, which can be organized into various topologies.
5. What properties are reasonable to infere a useful schema from 1.?

Future Perspective

Goals

What we'll *most probably* answer:

- What do we need to infere a schema?
- How is schema inference technically feasible?
- What do we need for the best possible classification?
- Which libraries are suitable for the construction of a prototype?
- What do developers need to create an application?

What we **can't** do:

- A productively usable tool for schema inference.
- Provide a maintainable tool.
- An additional tool for manual rectification.

References I

Adams, H., A. Tausz, and M. Vejdemo-Johansson (2014). “JavaPlex: A research software package for persistent (co) homology”. In: *International Congress on Mathematical Software*. Springer, pp. 129–136.

Bastian Rieck Max Hornung, E. G. (2016). *Aleph — A Library for Exploring Persistent Homology*.

Battle, L., R. Chang, and M. Stonebraker (2016). “Dynamic Prefetching of Data Tiles for Interactive Visualization”. In: *Proceedings of the 2016 International Conference on Management of Data*. Ed. by F. Özcan, G. Koutrika, and S. Madden. ACM, pp. 1363–1375.

Bauer, U., M. Kerber, J. Reininghaus, and H. Wagner (2017). “Phat–persistent homology algorithms toolbox”. In: *Journal of symbolic computation* 78, pp. 76–90.

Bowen, J. (2012). *Getting Started with Talend Open Studio for Data Integration*. Packt Publishing Ltd.

Budiu, M., R. Isaacs, D. Murray, G. D. Plotkin, P. Barham, S. Al-Kiswany, Y. Boshmaf, Q. Luo, and A. Andoni (2016). “Interacting with Large Distributed Datasets Using Sketch”. In: *16th Eurographics Symposium on Parallel Graphics and Visualization*. Ed. by E. Gobbetti and W. Bethel. Eurographics Association, pp. 31–43.

Chu, X., I. F. Ilyas, P. Papotti, and Y. Ye (2014). “RuleMiner: Data quality rules discovery”. In: *2014 IEEE 30th International Conference on Data Engineering*. IEEE, pp. 1222–1225.

References II

Cui, Z., S. Badam, M. Yalçin, and N. Elmqvist (2019). "DataSite: Proactive visual data exploration with computation of insight-based recommendations". In: *Information Visualization* 18.2.

Dasu, T., T. Johnson, S. Muthukrishnan, and V. Shkapenyuk (2002). "Mining database structure; or, how to build a data quality browser". In: *Proceedings of the 2002 ACM SIGMOD international conference on Management of data*, pp. 240–251.

Dhamdhere, K., K. S. McCurley, R. Nahmias, M. Sundararajan, and Q. Yan (2017). "Analyza: Exploring Data with Conversation". In: *Proceedings of the 22nd International Conference on Intelligent User Interfaces*. Ed. by G. A. Papadopoulos, T. Kuflik, F. Chen, C. Duarte, and W. Fu. ACM, pp. 493–504.

El-Hindi, M., Z. Zhao, C. Binnig, and T. Kraska (2016). "VisTrees: fast indexes for interactive data exploration". In: *Proceedings of the Workshop on Human-In-the-Loop Data Analytics*. Ed. by C. Binnig, A. Fekete, and A. Nandi. ACM, p. 5.

Feng, M., C. Deng, E. M. Peck, and L. Harrison (2017). "HindSight: Encouraging Exploration through Direct Encoding of Personal Interaction History". In: *IEEE Trans. Vis. Comput. Graph.* 23.1, pp. 351–360.

García, M. and B. Harmsen (2012). *Qlikview 11 for developers*. Packt Publishing Ltd.

Golab, L., H. Karloff, F. Korn, and D. Srivastava (2010). "Data auditor: Exploring data quality and semantics using pattern tableaux". In: *Proceedings of the VLDB Endowment* 3.1-2, pp. 1641–1644.

References III

Gratzl, S., N. Gehlenborg, A. Lex, H. Pfister, and M. Streit (2014). "Domino: Extracting, Comparing, and Manipulating Subsets Across Multiple Tabular Datasets". In: *IEEE Trans. Vis. Comput. Graph.* 20.12, pp. 2023–2032.

Gratzl, S., A. Lex, N. Gehlenborg, H. Pfister, and M. Streit (2013). "LineUp: Visual Analysis of Multi-Attribute Rankings". In: *IEEE Trans. Vis. Comput. Graph.* 19.12, pp. 2277–2286.

Gubanov, M. N., M. Stonebraker, and D. Bruckner (2014). "Text and structured data fusion in data tamer at scale". In: *IEEE 30th International Conference on Data Engineering, Chicago, ICDE 2014, IL, USA, March 31 - April 4, 2014*, pp. 1258–1261.

Hellerstein, J., C. Ré, F. Schoppmann, D. Z. Wang, E. Fratkin, A. Gorajek, K. S. Ng, C. Welton, X. Feng, K. Li, et al. (2012). "The MADlib analytics library or MAD skills, the SQL". In: *arXiv preprint arXiv:1208.4165.*

Iyer, G. R., S. Duttaduwarah, and A. Sharma (2017). "DataScope: Interactive visual exploratory dashboards for large multidimensional data". In: *IEEE Workshop on Visual Analytics in Healthcare*, pp. 17–23.

Javed, W. and N. Elmqvist (2013). "ExPlates: Spatializing Interactive Analysis to Scaffold Visual Exploration". In: *Comput. Graph. Forum* 32.3, pp. 441–450.

Kelly, J. E. (2015). "Computing, cognition and the future of knowing". In: *Whitepaper, IBM Reseach* 2.

References IV

Kohonen (2014). URL: https://pythonhosted.org/kohonen/.

Koytek, P., C. Perin, J. Vermeulen, E. André, and S. Carpendale (2018). “MyBrush: Brushing and Linking with Personal Agency”. In: *IEEE Trans. Vis. Comput. Graph.* 24.1, pp. 605–615.

Kraska, T. (2018). “Northstar: An Interactive Data Science System”. In: *Proc. VLDB Endow.* 11.12, pp. 2150–2164.

Law, P.-M., R. Basole, and Y. Wu (2019). “Duet: Helping Data Analysis Novices Conduct Pairwise Comparisons by Minimal Specification”. In: *IEEE Trans. Vis. Comput. Graph.* 25.1, pp. 427–437.

Lex, A., N. Gehlenborg, H. Strobelt, R. Vuillemot, and H. Pfister (2014). “UpSet: Visualization of Intersecting Sets”. In: *IEEE Trans. Vis. Comput. Graph.* 20.12, pp. 1983–1992.

Lin, H., S. Gao, D. Gotz, F. Du, J. He, and N. Cao (2018). “RCLens: Interactive Rare Category Exploration and Identification”. In: *IEEE Trans. Vis. Comput. Graph.* 24.7, pp. 2223–2237.

Liu, Z., B. Jiang, and J. Heer (2013). “*imMens*: Real-time Visual Querying of Big Data”. In: *Comput. Graph. Forum* 32.3, pp. 421–430.

Macke, S., Y. Zhang, S. Huang, and A. Parameswaran (2018). “Adaptive Sampling for Rapidly Matching Histograms”. In: *Proc. VLDB Endow.* 11.10, pp. 1262–1275.

Maria, C., J. Boissonnat, M. Glisse, and M. Yvinec (2014). “The Gudhi Library: Simplicial Complexes and Persistent Homology”. In: *Mathematical Software - ICMS 2014 - 4th International Congress, Seoul, South Korea, August 5-9, 2014. Proceedings*, pp. 167–174.

References V

Marr, B. (2017). *Data strategy: How to profit from a world of big data, analytics and the internet of things*. Kogan Page Publishers.

Mei, H., W. Chen, Y. Ma, H. Guan, and W. Hu (2018). "VisComposer: A Visual Programmable Composition Environment for Information Visualization". In: *Vis. Informatics* 2.1, pp. 71–81.

Morozov, D. (2007). *Dionysus, a C++ library for computing persistent homology*.

Murray, D. G. (2013). *Tableau your data!: fast and easy visual analysis with tableau software*. John Wiley & Sons.

Niederer, C., H. Stitz, R. Hourieh, F. Grassinger, W. Aigner, and M. Streit (2018). "TACO: Visualizing Changes in Tables Over Time". In: *IEEE Trans. Vis. Comput. Graph.* 24.1, pp. 677–686.

Papenbrock, T., T. Bergmann, M. Finke, J. Zwiener, and F. Naumann (2015). "Data profiling with metanome". In: *Proceedings of the VLDB Endowment* 8.12, pp. 1860–1863.

Pearson, P., D. Muellner, and G. Singh (2015). "TDAmapper: analyze high-dimensional data using discrete Morse theory". In: *R package version* 1.

Perin, C., P. Dragicevic, and J. Fekete (2014). "Revisiting Bertin Matrices: New Interactions for Crafting Tabular Visualizations". In: *IEEE Trans. Vis. Comput. Graph.* 20.12, pp. 2082–2091.

Powell, B. (2017). *Microsoft Power BI Cookbook: Creating Business Intelligence Solutions of Analytical Data Models, Reports, and Dashboards*. Packt Publishing Ltd.

References VI

Raman, V. and J. M. Hellerstein (2001). "Potter's wheel: An interactive data cleaning system". In: *VLDB*. Vol. 1, pp. 381–390.

Ren, D., M. Brehmer, B. Lee, T. Höllerer, and E. K. Choe (2017). "ChartAccent: Annotation for data-driven storytelling". In: *2017 IEEE Pacific Visualization Symposium*. Ed. by D. Weiskopf, Y. Wu, and T. Dwyer. IEEE Computer Society, pp. 230–239.

Sallam, R. L., J. Tapadinhas, J. Parenteau, D. Yuen, and B. Hostmann (2014). "Magic quadrant for business intelligence and analytics platforms". In: *Gartner RAS core research notes. Gartner, Stamford, CT*.

Satyanarayan, A. and J. Heer (2014a). "Authoring Narrative Visualizations with Ellipsis". In: *Comput. Graph. Forum* 33.3, pp. 361–370.

– (2014b). "Lyra: An Interactive Visualization Design Environment". In: *Comput. Graph. Forum* 33.3, pp. 351–360.

Siddiqui, T., A. Kim, J. Lee, K. Karahalios, and A. G. Parameswaran (2016). "Effortless Data Exploration with zenvisage: An Expressive and Interactive Visual Analytics System". In: *Proc. VLDB Endow.* 10.4, pp. 457–468.

Sisence (2013). URL: https://www.sisense.com/product/.

References VII

Srinivasan, A., S. M. Drucker, A. Endert, and J. T. Stasko (2019). "Augmenting Visualizations with Interactive Data Facts to Facilitate Interpretation and Communication". In: *IEEE Trans. Vis. Comput. Graph.* 25.1, pp. 672–681.

Stolper, C. D., A. Perer, and D. Gotz (2014). "Progressive Visual Analytics: User-Driven Visual Exploration of In-Progress Analytics". In: *IEEE Trans. Vis. Comput. Graph.* 20.12, pp. 1653–1662.

Tierny, J., G. Favelier, J. A. Levine, C. Gueunet, and M. Michaux (2017). "The topology toolkit". In: *IEEE Transactions on Visualization and Computer Graphics* 24.1, pp. 832–842.

Vartak, M., S. Rahman, S. Madden, A. G. Parameswaran, and N. Polyzotis (2015). "SEEDB: Efficient Data-Driven Visualization Recommendations to Support Visual Analytics". In: *Proc. VLDB Endow.* 8.13, pp. 2182–2193.

Wall, E., S. Das, R. Chawla, B. Kalidindi, E. T. Brown, and A. Endert (2018). "Podium: Ranking Data Using Mixed-Initiative Visual Analytics". In: *IEEE Trans. Vis. Comput. Graph.* 24.1, pp. 288–297.

Wang, Y., H. Zhang, H. Huang, X. Chen, Q. Yin, Z. Hou, D. Zhang, Q. Luo, and H. Qu. "InfoNice: Easy Creation of Information Graphics". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Ed. by R. L. Mandryk, M. Hancock, M. Perry, and A. L. Cox, p. 335.

References VIII

Wang, Z., N. Ferreira, Y. Wei, A. S. Bhaskar, and C. Scheidegger (2017). "Gaussian Cubes: Real-Time Modeling for Visual Exploration of Large Multidimensional Datasets". In: *IEEE Trans. Vis. Comput. Graph.* 23.1, pp. 681–690.

Wongsuphasawat, K., Z. Qu, D. Moritz, R. Chang, F. Ouk, A. Anand, J. D. Mackinlay, B. Howe, and J. Heer (2017). "Voyager 2: Augmenting Visual Analysis with Partial View Specifications". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Ed. by G. Mark, S. R. Fussell, C. Lampe, m. c. schraefel, J. P. Hourcade, C. Appert, and D. Wigdor. ACM, pp. 2648–2659.

Xia, J., W. Chen, Y. Hou, W. Hu, X. Huang, and D. S. Ebert (2016). "DimScanner: A relation-based visual exploration approach towards data dimension inspection". In: *2016 IEEE Conference on Visual Analytics Science and Technology*. Ed. by G. L. Andrienko, S. Liu, and J. T. Stasko. IEEE Computer Society, pp. 81–90.

Yalçin, M., N. Elmqvist, and B. Bederson (2018). "Keshif: Rapid and Expressive Tabular Data Exploration for Novices". In: *IEEE Trans. Vis. Comput. Graph.* 24.8, pp. 2339–2352.

Yu, B. and C. T. Silva (2017). "VisFlow - Web-based Visualization Framework for Tabular Data with a Subset Flow Model". In: *IEEE Trans. Vis. Comput. Graph.* 23.1, pp. 251–260.

Zgraggen, E., R. C. Zeleznik, and S. M. Drucker (2014). "PanoramicData: Data Analysis through Pen & Touch". In: *IEEE Trans. Vis. Comput. Graph.* 20.12, pp. 2112–2121.