# ADAPTIVE DATA VISUALIZATION TECHNIQUES FOR BUSINESS DATA ANALYTICS

# D09

IDEALVis Consortium

http://idealvis.inspirecenter.org/

European Union
European Regional Development Fund

Republic of Cyprus

Structural Funds
of the European Union in Cyprus

ΙΔΡΥΜΑ
ΕΡΕΥΝΑΣ ΚΑΙ
ΚΑΙΝΟΤΟΜΙΑΣ

# Executive Summary

This task presents an extensive background literature review that investigates research results and state-of-the-art tools and frameworks in the area of visual analytics and data visualizations content and cognition. Moreover, the literature review covers work done on state-of-the-art applications, new methodologies, theories, and techniques that will support the adaptation and personalization process, and the design and development of the IDEALVis platform. Subsequently, the information collected here, will guide the modification of current adaptation and personalization techniques as well as the creation of new processes, intelligent methods and interventions for enhancing and/or altering the data visualizations content, based on the human-centered user model. In addition, by investigating state-of-the-art tools, frameworks and techniques in the area of visual analytics and data visualizations, we verify that the global market does not offer products or services which are direct or indirect substitutes or are transforming to our targeted solution. Likewise, through this market analysis we clearly present a gap that exists in the leading data analysis and business intelligence platforms (namely visualization personalization), that IDEALVis promises to fulfil, through the use of new and existing techniques and methodologies based on research.

# Table of Contents

# List of Figures

# List of Tables

# 1 Introduction

In a modernized world that produces huge amounts of data daily, the need for techniques for making clear sense of the data is rapidly growing. Businesses use Data Analytics and Business Intelligence (DA / BI) platforms for analyzing their data and coming to beneficial decisions. The majority of such tools combine data analysis techniques with interactive information visualizations, for enhancing the decision maker's understanding of complex data for enabling effective decision support and decision making. While most modern tools can automatically calculate the "best suited" visualization based on the underlying data and the required task i.e. – view the weekly sales amount for two stores during the summer season (line chart) - view the total sales of two stores during the summer season (bar chart), they neglect individual differences and thus fail to deliver a personalized visualization experience. We argue that, there exists a number of influential human differences, such as the user's cognitive characteristics, personality traits, expertise and experience, which interfere with the processing and the subsequent understanding of visual information. Understanding which are those differences and their effect on visual information perception and processing is a vital requirement for moving towards personalizing visualization systems. A non-personalized visualization, that does not account for individual differences, can severely decrease the user's ability to understand the data, and thus result in poor decision making.

Section 2 of this literature review presents the value of information visualizations in a data driven society, by showing the volume of data growth through the past years. Section 3 focuses on the current data visualization techniques incorporated by DA / BI Platforms, and further explains why and how those visualization techniques are not personalized. Section 4 explores the importance of individual differences in information processing, along with Section 5 which specifically focuses on graph perception. Moreover, Section 6 describes how other systems including information visualizations have been adapted and personalized. Lastly, the above sections will help us draw a preliminary rationale for the visualization personalization techniques to be used in the current project.

# 2 The Gigantic, Non-Stop Data Generation

During the last years, the explosion in the volume of data produced has led to a new era of "Big Data" exploration and utilization, revolutionizing many businesses domains, not only with respect to their digital transformation, but also to the adoption of a new culture in decision making (Schrage, 2016). IBM states that, "Every day we create 2.5 quintillion bytes of data" (IBM, n.d.) – so much data that 90% of all the data in the word today has been created in the last two years alone, a statement supported by numerous studies (Domo, n.d.). More specifically, since 2016 there are six times more companies with over 1,000 terabytes of data, out of which 79% want to extract more value from the data and 70% need better analytics (Forrester, 2018). To address the data/analytics challenges, companies must continuously enhance their DA / BI Platforms with specialized modern capabilities (Sallam, et al., 2018). This can be achieved by adopting seamless data integration (Laney, et al., 2013) to compelling static or interactive visualizations – the latter quickly becoming a defining feature for effective visual-based data exploration (Richardson, et al., 2020). Combining existing data analysis techniques (e.g. data mining algorithms, predictive models) with interactive data visualizations can significantly improve the understanding of complex data (Liu, et al., 2017), leading to more effective decision making. To this end, modern DA / BI Platforms offer vast repositories of visual data analysis tools, techniques and myriads of customizable visualizations, of which a discussion takes place in the next section.

# 3 Data Visualization Techniques in Leading DA / BI Platforms

The demand for data analysis and business intelligence skills and solutions, has rapidly increased in the last 5 years, with an average rate of 25% to 50% per year. Today the market for intelligent data analysis and business decision support solutions is at its peak. The demand is driven by the business segment for solutions that would allow business experts, managers and other information workers, to take quicker and effective decisions, increasing both customer satisfaction and corporate revenue. Statistics show that nearly 50% of the enterprises have already either deployed a Big Data/BI solution or are in the process of doing so (Columbus, 2016). Therefore, it is recognized that efficient data exploration and recommendations on data visualizations provide great insights and value to businesses (Gentile, 2014), and multiple business solutions have developed to support the growth in interest. At a higher level, business visualization tools follow two main approaches. The first refers to expert users, mainly developers and data analysts, providing an interaction environment in the form of programming language libraries such as D3 and HighCharts; D3 is highly customisable and allows for the creation of new types of visualizations (D3JS, n.d.), whereas HighCharts is aimed more towards bootstrapping a common chart type that would be used by any developer (Highcharts, n.d.). On the other hand more user friendly solutions aim at users with no programming knowledge like Tableau that gives business experts the ability to explore data visually without the need for programming proficiency through providing an interactive user interface that allows for moving data through basic interfacing actions such as browsing selection and drag-and-drop (Tableau, n.d.). More on how platforms use such techniques in later sub sections.

In order to provide a wholistic picture of what data visualization techniques are used in leading DA / BI Platforms, we did a market research of the most prominent enterprise data analysis and visualization application software platforms. The challenge in this endeavor is that technology and marketed product offerings change at a very rapid rate, making it difficult to select the top tier companies for extracting their visualization techniques. Our first step in identifying which are the leading platforms, was to select a global market authority that ranks the different suites based on both market penetration as well as features and innovation. For this study, we have used the Magic Quadrant for Analytics and Business Intelligence Platforms Report (Richardson, et al., 2020).


*Figure 1- Magic Quadrant for Analytics and Business Intelligence Platforms 2020*

In this report, companies are scored and ranked against their ability to execute (which includes a combination of customer experience, product and service quality, ability to meet goals and market responsiveness / flexibility) and their completeness of vision (which includes a combination of innovation, market strategy and penetration and sales strategy). Companies which excel in both areas are mapped in the Leaders quadrant, whereas companies which have currently less capabilities but are targeting through innovation a higher market segment are mapped in the Visionaries' quadrant. Supplementary, another primary information source that was used to select the market predominant software and use them to compare and gauge the innovative nature of IDEALVis, was the blog article: "Is Big Data Still a Thing?" (Mattturck, 2016). Figure 2 provides a map of the application as extracted from the blog.

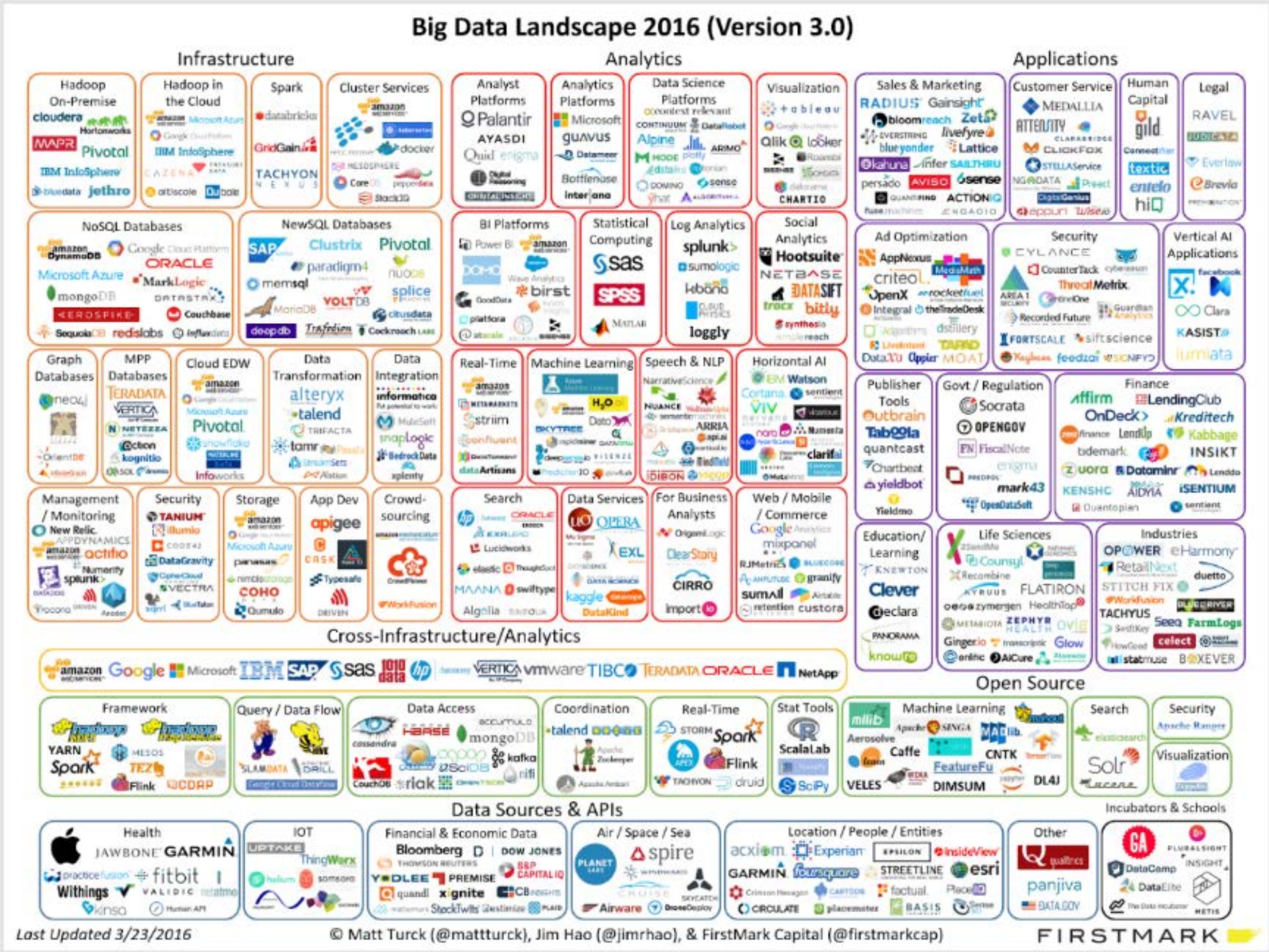


*Figure 2 - Big Data Landscape 2016 (Version 3.0)*

Our selection of platforms to investigate focused on all the companies in the Leader and Visionaries Quadrants, and a selected number of companies in the other Quadrants based on how related their offerings were to our research goals. A representation of the ranked companies in a graph is displayed in Figure 1 - extracted from the relevant report. The companies and their respective product selected for the analysis are presented along with the attributes analyzed for each, in Appendix 1. The analysis of the selected products and their attributes helped us understand what visual analytic capabilities they use. Moreover, the analysis of products was done to further determine if their offering provides or will provide in the near future, features in the area of cognitive data adaptive / personalised visualizations, to our understanding there is no current offering that is a direct or indirect substitute, or transforming to our targeted solution.

### *3.1 DA / BI Platforms Market Direction*

Prior to analyzing the data visualization techniques implemented by the (selected) most prominent vendors of DA / BI Platforms, it is important to understand the market direction those vendors are taking by examining some of the (market direction relevant) evaluation criteria used by the Magic Quadrant study. The criterion called Product Strategy (part of Completeness of Vision) examines whether vendors can keep up with trending features of the market, by measuring a vendor's Consumerization and Automation, two interconnected terms that represent, ease of use (as a matter of what analysis a planform does for the user, instead of how easy is for the user to perform analysis) and augmentation (use of machine learning and artificial intelligence for data preparation, insight generation and insight explanation) that acts as a path for automation. Moreover, as (Richardson, et al., 2020) states, augmented analytical capabilities are becoming key differentiators among platforms, with augmented analytics technology estimated to be ubiquitous by 2022".

With the above context we understand that the DA / BI market is moving towards a more automated and user-friendly platform conception, that will eventually enable non expert analysts to perform complicated analysis tasks.

## 3.2 DA / BI Platforms Visualization Techniques

Visual Analytics is a multidimensional concept that evolved through visualizations and algorithmic data analysis, that helps to address the information overload problem and discover knowledge in data by using visualizations. In visual analytics, visualizations are not used just as a results presentation tool, but instead as a tool that integrates the human cognition, perception abilities, and human intelligence into the data-analysis process to visually obtain explainable results, patterns and get insights from large data sets (Cui , 2019). As the DA / BI Platforms market is shifting towards a direction of automation and extended ease of use, visual analytics interaction and data visualization techniques in most platforms were also adapted into following the same path (Richardson, et al., 2020).

Automation, Customization, Summarization and Expandability are some key aspects that summarize the visualization techniques incorporated by DA / BI Platforms (Richardson, et al., 2020). In the following subsections, we expand each aspect by providing detail on how different vendors implement several visualization techniques for each of those aspects. Finally, we show that individual differences are a key aspect that is missing from how DA / BI Platforms are delivering visualizations.

### 3.2.1 AUTOMATION

A number of platforms can infer the type of the visualization to be loaded by inspecting the relationships of the selected data (TIBCO, n.d.) or the structure of the selected data (PowerBI, n.d.). Such fully automated approaches save the user a lot of time and minimize the chance of an error occurring, while also allowing the user (if desired) to then select a different visualization for the selected data (other than the automatically suggested one) from a list of visualizations compatible with the selected data. Other platforms like (MicroStrategy, n.d.) take a more guided approach to visualizing data, by allowing the user to preselect a desired visualization and then select specific types of (compatible) data to be placed on the visualization axes. Other platforms make use of natural language querying (NLQ), where a user can ask the platform for a visualization insight directly from the platform's interface (ThoughtSpot, n.d.) or by using a chatbot integration (Sisense, n.d.) that is communicating with the underlying platform. For example, a user can type in the search or in the chat "give me the sales of the last 2 years" and the system will reply with a bar chart illustrating the requested data.

### 3.2.2 CUSTOMIZATION

The majority of platforms allow users to create a dashboard from scratch using drag and drop, allowing for an easy experience when positioning components. A dashboard can be sliced in multiple areas that act as placeholders for the visualizations. Once a visualization is placed in a placeholder it can then be resized and further customized in multiple ways. Customizing a visualization goes beyond editing its visual elements i.e. size and color of bars in a bar chart or the thickness of a line in a line chart. Visualization backgrounds and color themes can be applied as well. Moreover titles, legends, tooltips and their font style or color, field formats i.e. decimal points or currency symbols are common visualization customization techniques found in the majority of DA / BI Platforms. Drilling up and down are also common techniques used to make navigating data inside a visualization flexible, this is done with linking, for example clicking on a country on a map visualization, loads another map visualization that contains the cities of the selected country. Finally, a number of platforms such as (Looker, n.d.) allow for combining chart types together for presenting more information and conveying a broader message. Combination of visualization types essentially allows users to display 2 visualizations in one, for example display a bar chart and a line chart that represents another data field, such visualizations are usually referred to as combo charts.

### 3.2.3 SUMMARIZATION

As a number of vendors are targeting a wider range of audiences i.e. not expert analysts a novel technique used to help the user better understand a visualization is natural language generation (NLG) (ThoughtSpot, n.d.) (OracleAnalyticsCloud, n.d.) (EinsteinAnalytics, n.d.). With NLG a platform autogenerates a narrative that summarizes the data displayed by the visualization for clearly passing the meaning to its viewer. Moreover, other than state of the art techniques such as NLG, vendors make sure to include the more basic techniques of simple summarization techniques such as aggregation, that in turn integrates with visualizations for conveying a clearer message i.e. grouping and displaying sales by state and by quarter on a map of the US (PowerBI, n.d.).

### 3.2.4 EXPANDABILITY

A number of visualizations exist for displaying a number of data types and most of the DA / BI Platforms by default provide more than the basic visualizations (bar charts, line chart, scatter plot, area chart, pie chart, map graph). Being able to expand the default set of visualizations provided by a platform though is necessity for many users. A number of vendors provide ways for visualization expandability, with two methods being the most popular. The first method is the use of third-party JavaScript Visualization Libraries and SDKs (YellowFin, n.d.) (Looker, n.d.) (MicroStrategy, n.d.), where users can write code and add their own custom visualizations on the platform. Finally, the last method of visualization expandability is done though vendor provided web stores or asset libraries (PowerBI, n.d.) (Sisense, n.d.) (Tableau, n.d.) where the users can download plugins and visualizations for expanding the default capabilities of the platform. Moreover, users can also publish a visualization they created in most of the online web stores for others to download and make use as well.

### 3.2.5 WHAT IS MISSING?

As the industry of DA / BI Platforms is shifting towards the engagement of non-specialized users that are not trained in advanced statistics or data science (Richardson, et al., 2020), to perform complex data analysis tasks; those users end up with a large number of tools and utilities that need to be orchestrated in order to make sense of the data and articulate their meaning using the most appropriate visualization; a cumbersome task even for experts. As of the above analyzation of visualization techniques, we can understand that the majority of tools can automatically generate multiple types of visualizations using a variety of data formats (Shneiderman, 1996), but with the human factors - which is one of the visual analytic facets (Kerren & Schreiber, 2012) - not being accounted for in the visualization process. While DA / BI Platforms fail to deliver a personalized visualization solution that accounts for human factors i.e. human cognition, and support the data exploration process, the analysis and understanding of data becomes demanding, time consuming, costly and even impossible (Liu, et al., 2014) since the potential non expert users (which DA / BI Platforms engage) are overloaded from the vast amount of visual information.

# 4 Individual Differences in Information Processing

Research in this field focuses on the following main categories of individual differences in information processing: a) cognitive styles (Koć-Januchta, et al., 2017) (Riding & Douglas, 1993) (Tsianos, et al., 2009) (Mawad, et al., 2015), which investigate how a user organizes and processes information; b) cognitive skills (Toker, et al., 2012) (Lallé, et al., 2017) ( Toker, et al., 2013), which investigate user cognitive characteristics that influence the user's effectiveness and satisfaction with a visualization; c) personality traits (Green & Fisher, 2010) (Ziemkiewicz, et al., 2011), which investigate how personality affects visualization interface interactions and visualization compatibility; and d) other characteristics (Toker, et al., 2012) (Lallé, et al., 2017) (Lee, et al., 2016), such as expertise, experience etc. In the following sections we explore literature that focuses on each of the abovementioned characteristics for better understanding their effect in information processing.

## *4.1 Cognitive Styles*

The term cognitive style was introduced by (Allport, 1937) and has been described as a person's typical or habitual mode of problem solving, thinking, perceiving and remembering. (Messick, 1984) further describes styles as consistent individual differences in ways of organizing and processing information and experience. Moreover, styles have been also identified as a bridge between cognition and personality, and a means for understanding, and improving educational achievement (Sternberg & Grigorenko, 1997). The cognitive style construct though has been also called "elusive" by (Riding & Cheema, 1991) whom has analyzed and grouped multiple cognitive styles into two principal styles based on their correlations. The two principal styles or "style dimensions" concluded from the study were namely Wholist-Analyst (individual processes information in wholes or in parts) and Verbalizer-Imager (individual represents information during thinking verbally or in images). The analysis done in this study resulted in the Cognitive Style Analysis (CSA) theory where an individual is classified on two independent scales Wholist-Analyst and Verbalizer-Imager.

We consider the proposed dimensions important to our work, since they are directly related to characteristics that influence visualization comprehension. Visualizations are made up of textual and visual elements, of which both are necessary for fully conveying the visualization's imprinted message. Moreover, those style dimensions can influence our model's personalization decisions in terms of, (1) amount of text or (2) structure to show to a particular user. A number of studies have examined the effects of those cognitive dimensions, others have focused on the Verbalizer-Imager dimension in learning and education (Riding & Douglas, 1993), (Tsianos, et al., 2009) and (Koć-Januchta, et al., 2017) while others focused on the Wholist-Analyst dimension in the consumer sector (Mawad, et al., 2015).

#### 4.1.1 VISUALIZER VERBALIZER DIMENSION

CSA has been used in studies like (Riding & Douglas, 1993), where students were classified as Verbalizers, Intermediates and Imagers. The study investigated the effect of text-plus-text versus text-plus-picture conditions, and the students' cognitive styles on learning performance. 59 students were randomly assigned to a learning condition and their task was to learn about car brake systems. Imager students performed better with the text-plus-picture condition and they were also more prone to using drawings for answering recall type questions. Findings of the study support that imagers represent information in a picture mode and their performance suffers when information presented is fully verbal. Another study that made use of CSA is (Tsianos, et al., 2009). Further on, eye-tracking was used in a web learning environment for better understanding how individuals structure information based on their cognitive style. In the study 21 participants were classified as imagers, verbalizers and intermediates and afterwards participated in an e-learning course about algorithms in computer science. Eye movement results proved that imagers focused more on images whereas verbalizers mostly on texts. Moreover, the study has validated the effect of style in information processing within the context of e-learning hypermedia, but no evidence was given whether those findings extend to commercial web-settings. Another study that utilized eye tracking was (Koć-Januchta, et al., 2017) for further inspecting how verbal and visual learners differ in their way of learning from texts and pictures. In contrast to the above studies, this study required the participants to learn about two topics (using text and images) of different nature (conceptual / mechanical knowledge) while their eye-gaze was recorded. Moreover, using more than one cognitive test 32 participants were classified as pure verbalizes or visualizers excluding intermediates. Verbalizers entered irrelevant areas of pictures sooner than visualizers, also visualizers shifted their point of focus from picture to picture more frequently than verbalizers, whereas verbalizers shifted their point of focus from text to text more frequently than visualizers. In general, regarding comprehension visualizers outperformed verbalizers. Visualizer-Verbalizer effects applied regardless of topic. Drawbacks of the study include that it did not take into consideration other cognitive measures that could influence results, and it examined participants that were either highly classified as visualizers or verbalizers where in reality most people have both styles (visual and verbal) to some extent.

#### 4.1.2 WHOLIST ANALYST DIMENSION

The Wholist Analyst style dimension has multiple terms that define it, one of the principal terms is Field Dependence Independence (FDI) (Riding & Cheema, 1991). The FDI term has been proven not to be an intellectual style, but instead a construct that represents an individual's ability in separating information from its contextual surroundings. A field independent individual has less difficulty in separating information from its surroundings, while a field dependent individual will be more likely affected by external visual cues (Zhang, 2004).

Using eye tracking devices (Mawad, et al., 2015) investigated how the FDI construct affects information processing on the decision of consumers when choosing between yogurt labels that contain textual and visual signs. 133 participants were classified as field independent or field dependent using the Group Embedded Figures Test (GEFT). Then participants had to make a choice between 16 pairs of labels containing varying information about (fat sugar, label background, brand popularity and the traffic light system). Field independent participants spent more time processing the label information before making a decision, moreover, they had higher fixations on areas of interest, also suggesting that they were able to sustain attention longer; a finding that is also supported by (Guisande, et al., 2007). An important finding is that the two groups performed similar processing on simple elements like images, but they significantly differed on how they processed more complex nutritional information (table of text), with field independent participants having higher fixations. Another related finding is that the two groups did not differ significantly on processing the visual representation of nutritional information (traffic light system), suggesting that simple visual representations of complex information can encourage field dependent users to engage in processing more complex information in alternative ways. From the above study we can conclude that cognitive styles, in this case FDI can affect a user's decision, since field dependent and field independent users differ in which attributes they choose to focus on.

### *4.2 Cognitive Skills Expertise and Experience*

A number of cognitive skills have been thoroughly investigated by researchers in the endeavor to broaden the understanding of how individual differences impact information processing. In this section we go through some of the literature that primarily focuses on the skills of perceptual speed, verbal working memory, visual working memory and spatial working memory.

(Toker, et al., 2012) investigated the effect of a user's perceptual speed, verbal and visual working memory and expertise on the effectiveness of bar graphs and radar graphs. The abovementioned characteristics were collected from 35 participants, which then had to go through 2 scenarios and answer questions based on data depicted on both visualizations. Both scenarios required participants to answer comparison questions with scenario 2 containing more complicated comparisons. Completion time for each task was collected (performance) and finally participants had to provide their graph expertise, ease of use (how easy is to understand the graph) and preference for each graph. Results for scenario 1 showed that bar graphs had faster completion times than radar graphs. Moreover, it was confirmed that perceptual speed is a relevant characteristic that affects completion times in both graphs. It was also noticed that as perceptual speed increases the completion time between the two graphs narrows. The perceptual speed effect was also visible in the results of scenario 2, and this gives more evidence for the effect being a valid factor that influences visual information perception. Preference and ease of use was found to be similar between the participants and this suggests that both graphs are easy to understand. Some interesting findings show that participants with higher visual working memory had a higher preference in radar graphs. Further on, participants with lower verbal working memory had a higher ease of use for bar graphs. Moving on, participants with higher radar expertise had a stronger preference for radar graphs and participants with higher bar graph expertise also had a higher rate on ease of use for radar graphs. From this study we observe that user characteristics affect a user's experience with data visualizations, specifically perceptual speed affects performance whereas visual working memory and verbal working memory influenced subjective preference and ease-of-use, respectively. A fact of discussion is that in scenario 2 authors could not detect a significant effect related to visualization type, either because of the training affect that took place in scenario 1, or because in more complicated comparison tasks, bar graphs are as good as radar graphs.

One study that investigates individual differences and goes beyond functional visualization tasks is (Lallé, et al., 2017). This study extends research on the impact of individual differences and proves that past results of individual differences to user performance and satisfaction apply to a real-world tool, specifically an application for supporting preferential choices in public engagement. Moreover, we see how individual differences impact user experience and decision quality, and the use of a map and a deviation chart in terms of gaze behavior. In the experiment 166 relevant participants had to explore 7 factors provided to compare transit alternatives and rank their top 5 factors by order of priority. Then, the participants rated transit alternatives by viewing a deviation chart and a map of the transit route while their eye movements were captured. Data collected from participants prior to the experiment included age, gender (related to map reading), visualization expertise, perceptual speed, visual working memory, verbal working memory, spatial working memory, visual scanning, LOC, N4C and visual literacy. Moreover, the study had 3 dependent variables, usefulness per visualization and interface, confidence of participants rating and decision coherence (this was calculated using the results from when the participants ranked the factors and rated the transit alternatives). Results of the study showed that higher self-reported expertise participants made more coherent choices than those with lower expertise, low spatial memory participants (those can retain and process less easily spatial information) found the chart less useful than participants with high spatial memory. Furthermore, high visual working memory participants tended to prefer the deviation chart over the map and the definition of this characteristics explain why (storage and manipulation capacity of shapes and colors of visual objects). Participants with low levels of spatial memory, visual scanning, perceptual speed and visualization literacy were at a disadvantage when comparing visualizations. Lastly results related to eye tracking metrics indicated that visual working memory impacts fixation rate and spatial memory impacts the number of fixations and the fixation rate.

A factor that varies between individuals and poses an important steppingstone in making sense of a visualization, is experience. This individual difference was further explored in (Lee, et al., 2016) where we see a qualitative study in which 13 novice participants try to make sense of three visualizations (parallel-coordinate plot, the chord diagram and tree map) that they had never encountered before. The purpose of the study was to investigate how novice users made sense of unfamiliar visualizations. Participants were shown the said visualizations while their thought process of trying to make sense of the data was captured through interviews. After the analysis of the data, researchers came up with a model that explains 5 cognitive activities (1 encountering visualization, 2 constructing a frame, 3 exploring visualization, 4 questioning the frame, and 5 floundering on visualization) that a user goes through in order to make sense of the unfamiliar visualizations. Patterns of how individuals navigate from one activity to another were also investigated. An interesting finding posed through transitions, is that novice users do not tend to revise their frame (explanatory internal structure) once it is constructed in their mind. Moreover, this finding suggests that the first impression a user gets for a visualization is vital for further understanding and for avoiding floundering. Further on, results explain how some of the participants developed a strategy for avoiding floundering. They achieved this by initially focusing on something that they could understand or were interested mostly, second by getting informative textual information from the visualization and lastly by constructing and comparing frames that could provide a possible explanation of visual objects. The findings of this research can help us understand how we can further adapt a visualization based on whether the user has seen it before. What the proposed model in the study lacks to account for is emotions aroused during sensemaking ( Germanakos, et al., 2008), personal interest about the content, and how these factors affect the sensemaking process.

( Toker, et al., 2013) further investigated the effect of cognitive abilities (perceptual speed, verbal working memory, visual working memory) on gaze behavior and how this effect is influenced by task difficulty and visualization type, aiming to better understand how specific characteristics influence the processing of visual information. In the study 35 participants completed cognitive tests and reported their expertise regarding the two visualizations (bar and radar graphs), and lastly, they performed 14 comparison tasks for each of two visualizations while 10 tasks were simpler than the rest. Moreover, participants also provided their confidence for each task they performed. Task difficulty was measured with principal component analysis by aggregating 4 measures (task completion time, standard deviation of completion time, average confidence, average deviation of confidence). 5 areas of interest for both visualizations were chosen, and eye tracking features were reduced to three families using principal component analysis (task level features, areas of interest proportionate features and areas of interest transitions). Mixed Model analysis was performed. Findings of the study explain that participants with higher perceptual speed had a higher fixation rate than lower perceptual speed users and lower fixation durations, thus they were scanning the screen more quickly with shorter and more consistent fixations while low perceptual speed participants spent more of their time in the legend and also transitioned to it more than high perpetual speed participants. For more difficult tasks all participants made more legend-related transitions (high perceptual speed made less) while for easy tasks participants made more label-related transitions (low perceptual speed made more). Moreover, it was found that the high area of interest had a high number of transitions by all participants in the radar graph, this effect was higher for low perceptual speed participants which suggests that this group is more affected by different ways of visualizing data. Further on, participants who had high verbal working memory, referred to the text question less often than low verbal working memory participants. Finally, the study did not report any findings regarding visual working memory, most likely because the graphs in the experiment were static and thus participants did not reach their maximum capacity for this visual capability.

### *4.3 Personality Traits*

A work that investigates personality traits is (Green & Fisher, 2010) and it studies the impact of locus of control (LOC), extraversion and neuroticism on visual analytics interface interaction and learning performance using two interactive visualizations (MapViewer, Gvis). The authors argue that personality factors affect the interaction outcomes, therefore if those personality differences are known, the interaction performance of the user can be predicted and subsequently used to create individualized interfaces. The above personality measures were collected for 106 participants, of which 50 participated in study 1 (in which they performed procedural tasks for finding information in both graphs), and 56 participated in study 2 (in which they had to demonstrate script learning). Completion time, insights generated (things learned) and visualization preference was also collected. Completion times were lower for MapViewer in both studies, but preference was higher for the Gvis. Findings suggest that participants with more internal LOC (believe they have control over personal life events) take less time in finding target information, compared to those that have more external LOC (believe that personal life events are not in their control). Neuroticism was negatively correlated with completion times in both interfaces and the faster the participant was the more extraverted. More errors were produced in the Gvis. LOC had an impact on insights generated with more insights generated by participants with more external LOC. Less extraverted participants reported more insights while more neurotic participants reported less than participants with lower neurotic scores. Authors noted that while participants with internal LOC had faster completion times, this is not the case when the task becomes complicated. One issue with the study is that findings do not differentiate between interface and interactive techniques. Lastly the study showed that personality factors predicted reported learning as well as performance and how this could be used to detect the user performance before he or she actually solves a task.

Locus of Control is also explored more closely by (Ziemkiewicz, et al., 2011) which tried to replicate the results of (Green & Fisher, 2010) but this time using a pure layout for displaying data. This study tried to assess the user's speed, accuracy, and preference with their locus of control on indentation metaphors and containment metaphors. The hypothesis explored is that LOC is affected more by layout in this sense than by visual encoding or interaction style. Based on (Green & Fisher, 2010) LOC influences an individual's use of a complex visualization system (includes visual encoding etc.), on the other hand (Ziemkiewicz, et al., 2011) suggests that the observed pattern may in fact be a correlation between LOC and visual layout. In (Green & Fisher, 2010) the 2 visualizations used had different visual encodings and interaction styles therefore (Ziemkiewicz, et al., 2011) proposes 4 visualizations that the variation between them is restricted to visual layout keeping interaction metaphor and visual encoding the same across all interfaces. 240 participants answered 2 questions (search and inferential) for each visualization. Preference of the user was also recorded for each visualization. Results show that participants with high internal LOC were slower than others when answering questions from the visualization that had a strong nested metaphor. On the other had high external LOC participants were more likely to perform quickly with the strong nested metaphor. Internal LOC participants were generally slower at answering questions correctly when compared to external LOC participants. Moreover, participants with a more external LOC were more accurate overall, while the other groups performed poorly with the visualization that had a strong nested metaphor. Also, internal LOC participants were much slower than others on inferential tasks but had the same speed when answering search tasks. From the above results we clearly see that a significant factor that interacts with LOC is layout (the way that visual elements are spatially arranged and presented), rather than interaction style or visual encoding. A very important result elicited from this study that is directly related to our personalization purposes is that we need to increase the amount of explicit structure for users that might have a more external LOC, and subsequently use a visualization with a simple spatial organization and minimal borders, outlines, and other grouping elements for high internal LOC users. This idea is based on the suggestion that LOC is predicting the degree to which a user will, automatically, prefer her own internal mental models (internal LOC) versus being willing to adapt to an external representation (external LOC).

# 5 Graph Perception

One work that helps us better understand how the structure of a visualization influences how we process it is (Ziemkiewicz & Kosara, 2008). A visualization consists of visual metaphors that help structure the information, i.e. containment metaphors thus the authors claim that the process of understanding a visualization involves an interaction with the visual metaphors and the user's internal knowledge representations. Moreover, the authors try to understand whether it makes sense to think of a visualization in terms of visual metaphors and also how these metaphors work to shape information. A motivation to this study is that several visualization evaluation studies have dissimilar results, due to the way they express visual metaphors in task questions to users. For example, two studies evaluating tree maps came to conflicting results because both of them were asking essentially the same questions to users, but using a different verbal metaphor, one tended to word questions in terms of depth and the other in terms of levels. Thus, this study further explored the effect of visual and verbal metaphors on the understanding of visualizations with an experiment where 33 participants had to answer comparison questions from data presented either on a tree map or a node-link diagram. Each participant had to answer 24 questions, of which half were worded to reflect a verbal metaphor incompatible to the visualization. A drawback to the study is that the visualizations used during the experiment were not interactive. Completion time and answer correctness of participants were analyzed and results indicated that the compatibility of visual and verbal metaphors can increase performance in visual processing, thus visual metaphors influence the representation of information in the mind, moreover, results proved that the visual metaphor affects how a user understands information from a visualization. Finally, the paper concludes that internalizing the visual metaphor is an important part of visualization perception.

A significant work in the field of graph perception was done by (Cleveland & McGill, 1984) where the authors try to take steps in establishing a scientific foundation for graphical methods used in data analysis and data presentation. The lack of this foundation motivates the authors to approach the science of graphs through human graphical perception and provide a set of guidelines that can inform the design of visualizations. With that in mind the authors make use of the graph perception theory, for inferring and ordering elementary perceptual tasks (position common scale, position non-aligned scale, length, direction, angle, area, volume, curvature, shading) on the basis of the accuracy with which people can extract quantitative information by using those perceptual tasks (the above elementary tasks are illustrated in the order inferred from the study). The authors demonstrate how each of those elementary tasks is used to extract quantitate information from a variety of graphs (plots, bar charts, pie charts, divided bar charts, statistical maps, curve-difference charts, cartesian graphs, triple scatterplots, volume charts and juxtaposed cartesian graphs) and then order those elementary tasks using theory, moreover, they validate their ordering using two experiments (position-length and position-angle). The main idea conveyed by the study is that, if two graphs present the same set of data, the graph that uses more elementary perceptual tasks that are ordered higher, will result in better and more correct perception of patterns and behavior of the data. To illustrate the above idea, the authors perform a “surgery” on the design of some existing graphs, by applying perceptual tasks as high in the order hierarchy as possible. One of the provided examples was the ordering of values of five categories (with similar values) on a pie chart. Value difference was not visible since all the slices seemed the same, therefore the authors using the elementary perceptual task ordering suggested a replacement for this graph. The replacement was a bar chart; therefore, we can see how the ability to perceive a pattern increased by changing the angle judgement to be a position judgement. A related research to (Cleveland & McGill, 1984) is (Cleveland & McGill, 1985) which further talks about the ordering of the visual metaphors done in their previous research and also explores numerus methods for analyzing quantitative scientific data.

A similar research to (Cleveland & McGill, 1984) that tried to provide a performance ranking of graphical encodings is the work of (Nowell, et al., 2002). Through an empirical evaluation of a digital library (Nowell, et al., 2002) suggested that a graphical encoding’s effectiveness is indicated better by the perceptual task performed by the user when processing an information visualization, rather by the data being represented by the visualization. The Graphic View used in the experiment was the search result display of the digital library called Envision, that could return relevant documents when issued a textual search query. The search result display was a scatterplot like representation, that illustrated document publication year on the X axis and index terms on the Y axis. Moreover, documents were presented inside the visualization with each document having a varied graphical encoding depending on its properties i.e. document’s type and document’s search relevance. The experiment made use of the result display explained above, for investigating the effectiveness of three graphical devices / encodings (color, size and shape) when presenting nominal (document type) and quantitative (document relevance) data about the documents in a result. 20 participants had to perform visual search and identification tasks while viewing result sets containing documents encoded with different conditions. Error rate and time for task completion were the dependent variables of the study. Results for time to task completion and error rate were calculated separately for document type (nominal) conditions and separately for document relevance (quantitative) conditions, resulting into 2 rankings for codes conveying nominal data and 2 rankings for codes conveying quantitative data. The first 2 rankings for conveying nominal data ranked color first, size second and shape third according to time to task completion and color first, shape second and size third according to error rate. Moreover, the next 2 rankings for conveying qualitative data ranked color first, shape second and size third according to time to task completion and color first, shape second and size third according to error rate. For drawing conclusions, the produced rankings were compared to other studies that ranked the same graphical devices. While the ranking of graphical devices for nominal data was similar with rankings found in other studies, the ranking of graphical devices for qualitative data was dissimilar from other studies such as (Cleveland & McGill, 1984) (Cleveland & McGill, 1985). This dissimilarity makes up the authors rationale, that the effectiveness of a graphical device differs among studies, due to the nature of tasks on which the ranking is based. Moving on, the authors suggest that designers should not pick a ranking that is solely based on the type of data, but one that is instead based on the user’s task and the appropriate measure of effectiveness, i.e. error rate or completion time. For when counting identification tasks are performed the ranking of the current study is appropriate, but when graphical perception tasks i.e. extraction of quantitative data or numerical comparisons between graphical objects the ranking of (Cleveland & McGill, 1984) (Cleveland & McGill, 1985) is suggested as more appropriate. Finally, further research on graphical devices such as letter, digits, flash rate, texture etc. is proposed and the use of color along with texture is encouraged for color impaired users.

Color has been the winner in multiple graphical encoding rankings as seen in (Nowell, et al., 2002). The power of color has been further leveraged by researchers (Hagh-Shenas & Interrante, 2005) in multivariate visualizations. (Hagh-Shenas & Interrante, 2005) demonstrated existing techniques of using color and texture as means of conveying multiple variables at a single spatial location within a visualization i.e. *color compositing* and *color weaving*. Further on, through the use of texture perception and a number of texture metrics such as directionality, coarseness and contrast the authors devised a new technique that given an arbitrary texture can automatically interweave colors in it. Compared to previous approaches such as compositing, this approach can allow for a larger number of colors to be represented and can work with any given texture, instead of being limited to Perlin noise textures as it is in the case of the color weaving technique. The proposed automatic texture coloring technique consists of 5 phases, (1) decomposing a given texture in multiple components based on the orientation of the texture's directional elements and differentiating texture foreground and background, (2) applying *steerable-pyramid* filters on texture patters, (3) removing small unreliable blobs from each of the directions, (4) using a filter to smooth the segmentation result and finally (5) the colorization phase. At the colorization phase the intersection of texture blobs with each of directions is colored using a data variable that is to be represented by that region. By applying the above technique, the authors used 2 textures for coloring the regions of a map visualization using 5 variables, one variable mapped by the texture type and the other four mapped by the presence or absence of colors following pattern directions i.e. vertical, horizontal or diagonal. Further on, it has to be noted that before applying this technique a close inspection on the texture pattern to be used is needed to make sure that the pattern's spatial frequency matches the number of variables that are to be visualized. Finally, the proposed approach has demonstrated that the use of color and texture together can help increase the number of variables visualized, especially on map like visualizations but without providing any evidence of how effective their produced visualization was.

(Cahill & Carter, Jr., 1976) focuses on better understanding how the number of colors in a color code affect the visual search tasks completion time. The authors noted the effectiveness of color in visual search tasks and further discussed that the two parameters of constructing a color code for a visual display, are the number of colors to be used (code size) and the total items presented on the display (density). The study's rationale is based on the fact that there should exist a minimum value for an intermediate degree of coding, since when no color is used or when all data points have a unique color (two extremes), the user has to go through all items since no color grouping exists to facilitate the visual elimination of unwanted values. The experiment was done with 20 participants that had to go through 50 trials each, trying to detect a given number of a given color on a display. The independent variables were density and number of colors used, and the dependent variable was search time. The displays for the 50 trials were constructed using 10 distinct colors applied on each of 5 levels of density from 10 to 50 data items (increasing by 10).

The experiment resulted in 1000 results and the following conclusions were drawn. As proven in previous research the effect of density was linear since the mean search time increased as the density of items on the display increased. Moreover, the effect of number of colors helped in decreasing search time as more colors were added, until a point where search time started increasing again as more colors were used even if the number of items per color was kept constant. Further on, the relation of number of colors and density to search time, was also examined and until 7 colors the search times were decreased in a variable manner in all different display densities. While previous studies showed that in visual search tasks a gain in efficiency (less time) can be achieved with codes of up to 5 colors, the current study suggests that with an appropriate selection of discriminable colors a code of more than 5 colors can be also considered safe. The suggested color code size for 10 to 20 items on a display is up to 10 colors and for a larger set of items (denser) even 8 or 9 colors could be used. Finally, for displays of more than 50 items the authors suggest that a larger code size won't be needed, and also note that if a combination of density / color reaches the point where 5 items are illustrated per color category, this might as well remove the benefit of using color to improve visual search efficiency.

Graph perception is largely affected by the task performed by the user. For example, the effectiveness of a specific visual encoding such as color might be different when the user is performing a search task and different when performing a filtering task. For better understanding the analytical tasks performed by a user when processing a visualization, the study of (Amar, et al., 2005) was considered. This study describes a new taxonomy of information visualization tasks that in contrast to other taxonomies, focuses on capturing the user's analytic activity. As further explained the set of low-level analytical tasks deduced from this study largely capture people's activities while employing information visualization tools for making sense of data. In the study, students had to go through data sets of different domains and generate data analysis questions while evaluating how well each question could be answered using a number of commercial visualization tools. After collecting 200 questions from students, the authors applied an affinity diagramming approach for grouping similar questions and extracting the knowledge goal of each group. The analysis of the 200 questions resulted in 10 low-level analysis tasks. The analytical tasks are Retrieve Value, Filter, Compute Derived Value, Find Extremum, Sort, Determine Range, Characterize Distribution, Find Anomalies, Cluster and Correlate. Since the tasks described in this research are expected to emerge as a user is processing an information visualization system for capturing some higher-level knowledge, the proposed tasks could be used as means for an informal evaluation of a new visualization system or technique as previously seen in (Toker, et al., 2012). The only concern with this study is the fact that the analytical tasks captured by the questions might be influenced by the students' previous course knowledge or even influenced by the visualization tools employed in the study. This concern though does not eliminate the results of the study since the resulting taxonomy had similarities to other taxonomies.

In a review of five books on graphs and charts (Kosslyn, 1985), the human visual information processing system was partitioned in three general phases, with the first phase being the process of getting visual information into the system itself, by detecting edges and isolating regions of a stimulus. The said phase was further split into factors that affect how well the said phase is executed. Here we discuss some of those factors and further discuss parts of the reviewed books that aim to deal with them. A factor called Processing Priorities describes how parts of the stimulus are by default detected easier than others. A simple example of this effect is shown when we tend to detect a bold line earlier than a lighter line. As mentioned in (Kosslyn, 1985) the book of (Tufte, 2001) provides a recommendation of how the design of the chart can encourage the user to perceive what matters first i.e. in a correct priority by increasing the amount of "data ink" (main area where data is visualized) and erasing as much "nondata ink" as possible assuming that a good reason exists. This idea was further judged as "extreme" by (Kosslyn, 1985) as in reality most "nondata" items of a chart i.e. title, are usually important to the user for conveying the underlying chart data. Processing Priorities and the suggestion of (Tufte, 2001) though, can help us think of other techniques of guiding the user to perceive elements with a given priority, such techniques can for example include (1) showing the title of the chart before the actual chart becomes visible to the user or (2) instead of totally removing an element i.e. axis lines, we could partially decrease its opacity to better direct the processing prioritization. A second factor found under the first phase of the human visual information processing system is what the reviewer calls Adequate Discriminability, which is described as a matter of how distinguishable a mark i.e. a letter or a point, is from the whole i.e. visualization, and its neighbor marks, as to the point that the mark is clearly noticeable and distinguishable. A related metric to discriminability is the minimum size a certain type of mark can take up to the point it is no more consciously visible. This metric is also known as the "absolute threshold" and for some types of marks there exists a calculation of this metric. (SMITH, 1979) for instance, examined the effects of letter size on legibility on computer screens, a variable that we can directly use and manipulate, since letters make up a lot of parts of visualizations i.e. legends and titles.

( Dimara, et al., 2016) investigated an important perceptual effect that explains how decision making through the use of visualizations can be affected from a cognitive bias called the attraction effect. The attraction effect essentially affects the decision-making process by driving people to favor an option (target) of which there exists a similar but slightly inferior alternative (decoy), thus performing an irrational decision. The motivation to the study is that the attraction effect was previously found effective on simple presentation formats such as tables and pictures, but its effects were not investigated on visualizations. In a first experiment the authors were able to prove that the attraction effect extends on scatter plots (when a third point is added) by replicating an older experiment done using a table. The experiment required participants to select between gyms that varied in cleanness and variety. Each participant viewed a single scatterplot (this experiment also had a table representation of the gyms, that we are not mentioning here) and had to select from it the preferred gym. Gyms were presented to each participant on one of three possible scatterplots (1) showing just two points representing the two target gyms, (2) showing three points representing the two target gyms and one gym acting as a decoy to the first target gym, (3) showing three points representing the two target gyms and one gym acting as a decoy to the second target gym. Moreover, the authors performed a second experiment since the first one only used three data points, which certainty does not capture most real-world decision tasks where visualizations are used. Using a new design, the authors replicated a previous study by using scatter plots, where participants had to make 20 decisions by choosing between lottery tickets that competed by win price and win probability. Each scatterplot used two non-dominated options (targets) plus several "distractors" (irrelevant data that cannot affect the decision) and multiple decoys, instead of one. Results concluded that when people explore choice alternatives on scatter plots, the number and position of inferior or irrelevant choices can influence their decision. While this paper did not investigate the origins of why the attraction effect is visible on scatter plots and its debiasing, its certainty the first InfoVis study on the attraction effect.

A graph is made up of multiple elements and reference structures or visual metadata that guide the user into better understanding the underlying data. (Stone & Bartram, 2009) explains that such visual metadata are important, but the same time need not be obtrusive in a way that clutters the visual information. The work of (Stone & Bartram, 2009) though a number of experiments, presents an effective range in terms of transparency that can be used on grids over scatterplot data. Two experiments were performed, where participants had to view scatterplots with varying background color, lightness and varying density of plotted points. For each scatter plot the participants had to perform 2 tasks, (1) adjust the grid's transparency until it became usably perceptible without being unnoticeable and (2) adjust the grid's transparency up to the point that it would be obvious but not intrusive. Experiment 1 and 2 were similar, in 1, participants manipulated a dark grid over a white plot and in 2, participants manipulated a white grid over a dark plot. Results of the study indicated a usable range defined by alpha for grids; the range suggests that for images that are not very dense a light grid can be created using a value of alpha around 0.1. Moreover, results indicate that a maximum grid alpha of 0.2 will be suitable or "not bad" for all other cases.

# 6 User Adaptive and Personalized Systems

This section is focused on the adaptation side of adaptive interactive systems. Specifically, the analysis specifies which visible aspects of a user interface should be adapted and how, what adaptation mechanisms should be implemented, how should the system's content and functionality should be structured and prepared for input to the adaptation mechanism, and how the adaptation effects on the user interface should be communicated. Accordingly, the section discusses the state of the art in adaptive user interfaces (i.e., various adaptation ways and effects), and adaptation mechanisms in various domains; at first discussed intentionally beyond the scope of the data visualizations area in an attempt to inclusively consider the outcome and lessons learned for a more optimized solution. Moreover, after the general discussion of adaptation in various domains, we turn the focus on works specifically related to visualization adaptation techniques and systems, visual guidance and the indirect collection of user characteristics for adaptation purposes (mostly focusing on visualization adaptation techniques).
In a second phase we discuss about some existing personalized visualization systems and various other non-visualization systems that were personalized, for better understanding the personalization landscape in other areas as well, for enhancing the choices to be made in our personalized system.

## 6.1 Adaptation Mechanisms

Adaptation mechanisms apply specific algorithms that decide what adaptation will be performed on the content and functionality of the system. Various approaches have been proposed in the literature, including among others user customization, rule-based, content-based and collaborative mechanisms.

### 6.1.1 USER CUSTOMIZATION

*User customization* provides a mechanism that allows users to construct a custom interface representation based on their own preferences. Once the user has entered this information, a matching process is used to find items that meet the specified criteria and display them to the user. The system in this case is not considered adaptive, but rather adaptable because it is explicitly configured by the user how to adapt its content and functionality. (Yen & Acay, 2009), for example, proposed a novel idea for adaptation of the user interface for complex supervisory tasks. An adaptive interface can be controlled by its user in the following ways (Horvitz, 1999) (Kühme, 1993) (Keeble & Macredie, 2000) (Oppermann, 1994).

1. Providing means to activate and deactivate adaptation partially or completely.
2. Providing means to set parameters in the adaptation algorithm.
3. Giving control over the use of behavior records and their evaluation (control over privacy).
4. Offering the adaptation in the form of a proposal (the user can accept or reject the adaptation).
5. Providing means to review and manage completed adaptations (the user can save/load previous adaptations).
6. Providing information on the effects of the adaptation.
7. Providing information on the rationale of the adaptation (transparency or predictability).

(Wang, et al., 2010), for example, described a framework for collaborative tagging social media systems, which allows users to annotate the resulting user-generated content, and enables effective retrieval of otherwise unstructured data. The personalized environment developed would be especially appropriate for the following tasks: collaborative tagging, collaborative browsing and collaborative search.

### 6.1.2 RULE-BASED MECHANISMS

Rule-based mechanisms refer to the process of producing high-level information from a set of low-level metrics, related to both static and dynamic user context information. Bearing in mind that the dynamic part of the context data model can be updated in real time it becomes obvious that reasoning capabilities supported provide an added value supporting users in different tasks. Such rules can initiate automated system actions or compare predictive user interaction models with actual user interaction data gathered in real time, providing thus valuable insights related to the current user goals and efficiency of interactions. For example, an online banking system may contain a rule "If ([USER].logged=False and [USER].loginattempts.count>2) then [UIOBJECT.LiveSupport.show=True]", which indicates that the system should automatically offer a live customer support option to users who could not succeed to login in the system after trying to login for more than two times. Based on another usage scenario such a rule-based adaptation mechanism could extremely increase usable security by offering a live customer support option to users whose e-Banking web accounts were locked due to numerous unsuccessfully login attempts. A detailed analysis and comparison of rule-based mechanisms can be found in (Smyth, 2007).

### 6.1.3 CONTENT-BASED MECHANISMS

Content-based mechanisms suggest labelling of links by analyzing the content of pages. A typical content-based mechanism includes the following steps: i) pre-fetch the content behind the links of the current page, ii) parse the pre-fetched pages to create a weighted keyword vector of each page, iii) compare the weighted keyword vector of each page with the user's preferences, that are also usually represented using a weighted keyword vector, iv) suggest pages whose keyword vectors are the same with the user's preferences. FishWrap (Chesnais, et al., 1995), for example, was one of the first prototypes of personalized newspapers using profiles of individual members of the MIT community. The system provided general news about the world and the university community. The user profile was developed by asking the user three questions: origin, affiliation in MIT and major interests and by recording user navigation. Additionally, the user could update their profile.

In web sites such as (Yahoo, n.d.), and (MSN, n.d.), the user typically selects categories of interest and the page is built on-the-fly to match the available content to his or her preferences. The content categories are usually quite broad, and the personalization lacks dynamic updating of user interests over time, and all changes are made manually. Consequently, users receive information on out-of-date categories until they update their fields of interest. This strategy consists in suggesting items similar to others that gained the target user's interest in the past (BRIDGE, et al., 2006), which is quite simple to implement. However, the recommendations tend to be repetitive for considering that a user will always appreciate the same kind of content. This overspecialization may not pose a problem with users who want to remain informed on specific topics (e.g. people with chronic diseases), but it does so in general.

### 6.1.4 COLLABORATIVE MECHANISMS

In response to the problem of overspecialization, researchers came up with collaborative filtering to consider the success of the recommendations previously made to users with similar interests (the neighbors of the target user) (Pazzani, 1999). This approach solves the lack of diversity but works poorly with users (the gray sheep) whose preferences or needs are dissimilar to those of the majority. Collaborative mechanisms exploit the social process of people of recommending something they have experienced with (e.g., read a book, watched a movie, etc.) to other people. Collaborative mechanisms are based on the assumption that if users X and Y rate n items similarly, or have similar behaviours (e.g., buying, watching), hence will have similar interests. Adaptive interactive systems utilize collaborative mechanisms to provide navigation support by recommending links of interest to the user based on earlier expressed ratings or navigation behaviour of similar users. (Amazon, n.d.) is largely based on this method, where a user's past shopping history is used to make recommendations for new products.

(Das, et al., 2007) described an approach to collaborative filtering for generating personalized recommendations for users of (Googlenews, n.d.). The site is not an online version of a traditional printed newspaper; but rather a collection of the most visited news article on the web. The user can change or delete the layout of topics and can state a number of keywords he or she would like to have in an article. (Aggarwal & Yu, 2002) describe a system for personalizing web portals containing news feed services. The system employs collaborative filtering techniques, and the personalization is achieved by both the user entering explicit information and by implicit input. ANATAGONOMY (Kamba, et al., 1997) personalizes web pages by monitoring user operations on

articles and creating user profiles based on both explicit and implicit feedback from the user. The system uses both content based and collaborative filtering techniques.
The most recent strategy is item-based collaborative filtering, which consists in recommending items related to others that the target user liked in the past, considering two items related when users who like the one tend to like the other as well (Sarwar, et al., 2001). This approach still faces several problems that were also apparent with collaborative filtering. One of those problems is sparsity, implying that when the number of items available to recommend is high (as it happens in many domains of recommender systems application nowadays), it is difficult to find users with similar valuations for common subsets. Another important drawback is that of latency, related to the inability to recommend recently added items, as long as there are no user ratings available for them. (Nores, et al., 2011) presented a new strategy, called property-based collaborative filtering in the context of health-aware recommender systems, as a means to tackle the aforementioned problems in general settings. This approach depends on having a semantic characterization of the items that may be recommended, which is not necessarily true for other mechanisms of adaptation (see also (Blanco-Fernández, et al., 2011)).

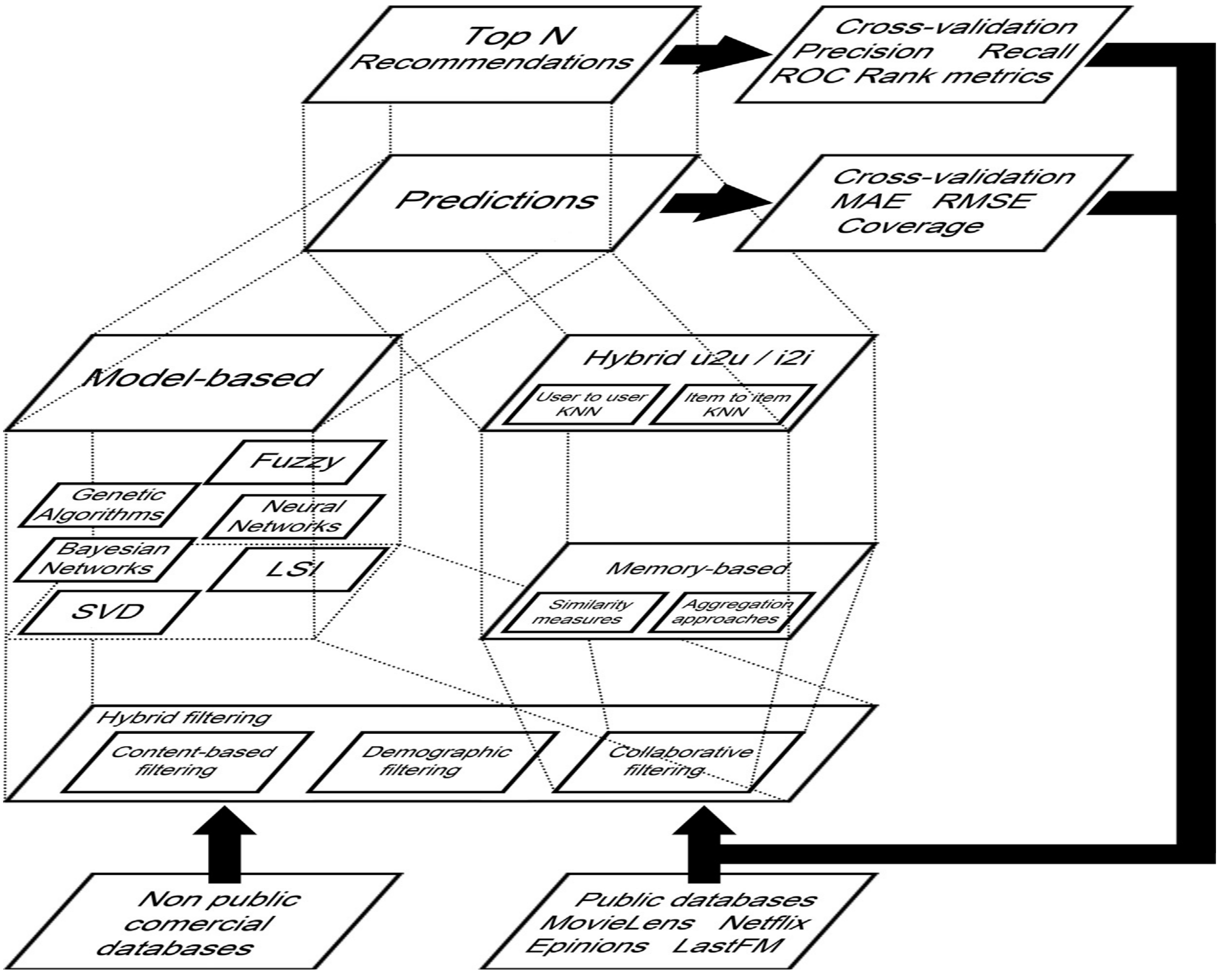


*Figure 3 - Traditional models of recommendations and their relationships (Bobadilla et al., 2013)*

Bobadilla et al. (2013) provides a detailed overview of the area of recommender systems in Figure 3 arguing that currently these systems may incorporate user social information (friends, followers, trusted users). (Bobadilla, et al., 2013) argues that in the future, systems will use implicit, local and personal information from the Internet of things/integrated devices on the Internet (e.g. location information, data from devices and sensors, real-time signals, weather parameters).

## 6.2 *Adaptation Effects*

Adaptation effects can include special navigational tools such as table of contents, index, maps and recommendations that could be used to navigate users to all accessible pages that can be adapted, here are the page (content-level adaptation) and the appearance and behavior of the links (link-level adaptation). In adaptive hypermedia literature they are referred respectively as adaptive presentation and adaptive navigation support. *Adaptive Presentation* is to adapt the content of a hypermedia page to the user's goals, knowledge and other information stored in the user model. There could be multiple reasons to use adaptive presentation. Two typical cases in the area of education are comparative explanations and explanation variants. The idea of comparative explanations is to connect new content to the existing knowledge of the learner. *Adaptive Navigation* support is to help users to find their paths in hyperspace by adapting link presentation to the goals, knowledge, and other characteristics of an individual user.

A good design practice aims to establish a common ground among designers and users related to the aspects of user-system interaction by formalizing the information architecture of the interactive system and specifying the interaction flow for accomplishing specific tasks. A well-used and simple approach to modelling interactive systems is to analyze the user actions in several levels of abstractions and identify on each level the most appropriate terminology, content presentation and interaction flow. The high-level architecture of modelling interactive systems is illustrated in Figure 4.

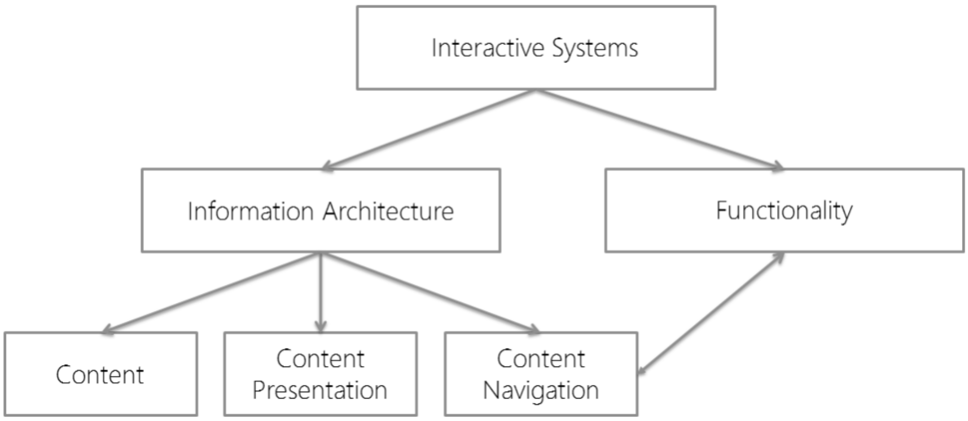


*Figure 4 - High-level Architecture of Interactive Systems*

An important adaptation issue in adaptive interactive systems is which visible features of the system can be adapted by a particular technique. According to (Brusilovsky , 2001), there exists a number of ways to adapt hypermedia. These are classified under two main classes of adaptation technologies; content-level adaptation, called adaptive presentation and link-level adaptation, called adaptive navigation support.

Adaptive presentation relates to the adaptation of hypermedia elements inside nodes, and adaptive navigation support relates to the adaptation of links inside nodes, indexes and maps. These are discussed next.

### 6.2.1 ADAPTIVE (CONTENT) PRESENTATION

Adaptive presentation relates to the adaptation of hypermedia elements inside nodes. The idea behind adaptive presentation is to adapt the information elements (or content) inside a node accessed by a particular user to the needs and preferences of that user. Adapting the presentation of content within a node is most often performed as a manipulation of fragments. Such manipulations aim to provide prerequisite, additional or comparative explanations. For example, additional information can be shown for users with a specific state of knowledge to provide missing prerequisite knowledge, additional details, or a comparison with a previously known concept.

Techniques that are used to provide adaptive presentation include: i) inserting/removing relevant to the user fragments, ii) expanding/collapsing content fragments (e.g., expand additional explanations to novice users), iii) altering content fragments (e.g., present a diagrammatical representation of a concept to an Imager cognitive style use ( Germanakos, et al., 2008)), and iv) sorting content fragments (e.g., some users may prefer to see an example before a definition, while others prefer it the other way around).

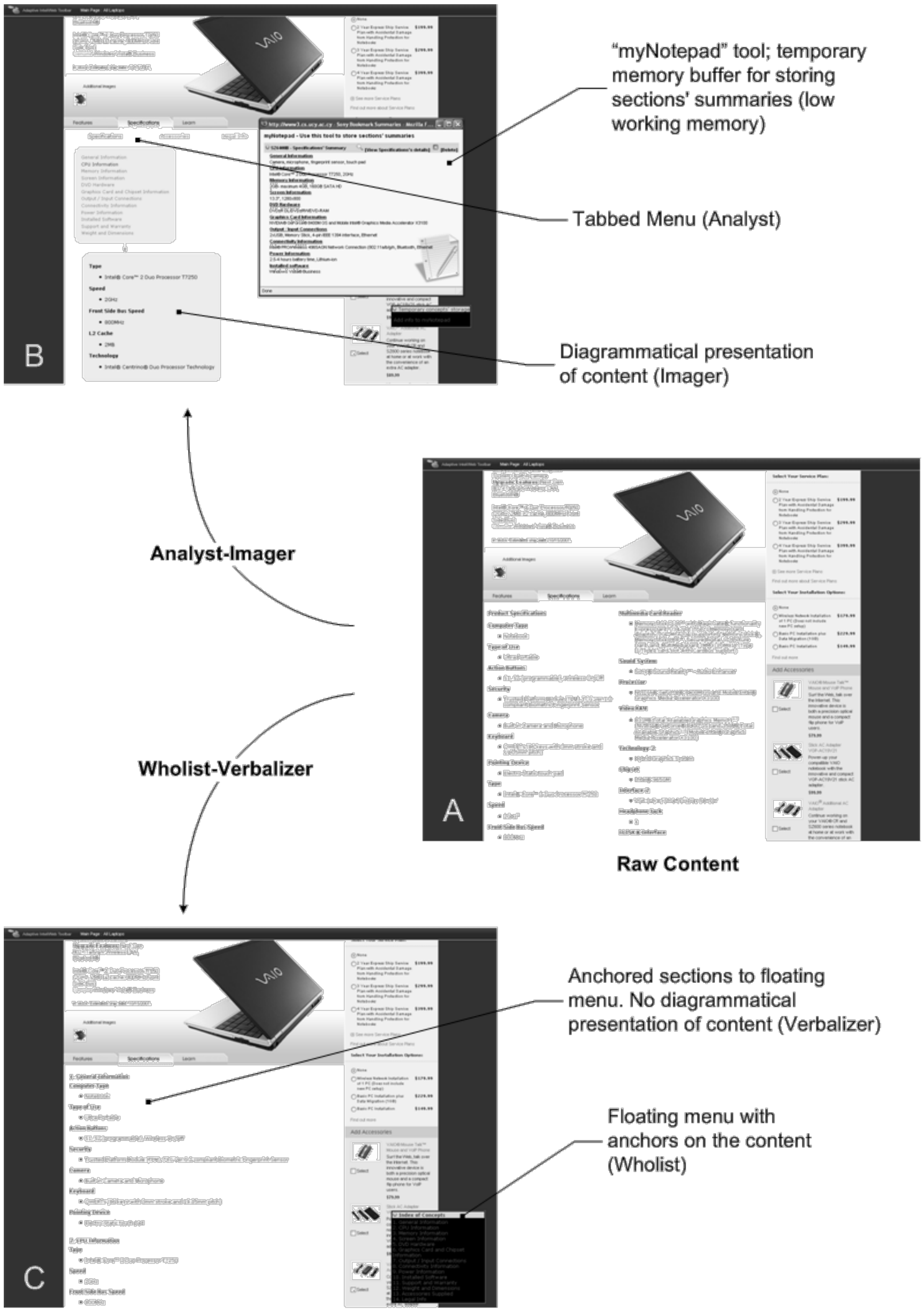


*Figure 5 - Content Adaptation based on Cognitive Styles of Users*

Figure 5 illustrates an example of content adaptation utilized in a previous study of the authors ( Germanakos, et al., 2008) where users with different cognitive typologies (i.e., Verbalizer, Imager, Intermediate) were provided with different content fragment variations, i.e., users belonging to the Verbalizer class (that process textual content efficiently) were presented with more textual content, whereas users belonging to the Imager class (that process graphical content efficiently) were presented with more graphical content. Furthermore, this study provided adaptive navigation support based on other cognitive factors (i.e., Wholist-Analyst) ( Germanakos, et al., 2008) that affect navigation behaviour of users in interactive systems.

We also provide an example of the Pandora music recommender system in Figure 6, where the user interacts with the system with the goal of finding a music item, and the system recommends items based on what it has learned about the user's interests.


*Figure 6 - Pandora music recommender example screenshot*

(Park & Han, 2011), for example, proposed a method of coupling adaptable and adaptive approaches to the design of menus. The proposed complementary menu types incorporate both adaptability and adaptivity by dividing and allocating menu adaptation roles to the user and the system. The results showed that adaptable and adaptive menus were superior to the traditional one in terms of both performance and user satisfaction. Specifically, providing system support to the adaptable menu not only increased the users' perception of the efficiency of selection, but also reduced the menu adaptation time. (Park & Han, 2011) suggested the possibility of designing adaptive web interfaces with user control (partly adaptable), which may provide additional advantages, such as psychologically increasing user control of the interaction, and requiring less effort for adaptation to his/her needs. Recently, (Kardaras, et al., 2013) applied Fuzzy logic techniques (Delphi method and Cognitive Maps) to content presentation and media adaptation on a tourism web site prototype. This research highlighted service features that are most preferred by users and ways to adapt presentation media and layout based on user preferences.

### 6.2.2 ADAPTIVE NAVIGATION SUPPORT

Adaptive navigation support relates to the adaptation of links inside nodes. This kind of adaptation supports user navigation in an interactive system by adapting to the goals, preferences and knowledge of the individual user. The core idea behind this kind of adaptation is to adapt the presentation of hyperlinks/functionality within a node. Adaptive navigation support can be achieved by: i) guiding the user in the system by suggesting the "next best" node to visit according to the user's goals, preferences and knowledge, ii) prioritizing links that are relevant to the user closest to the top, iii) by hiding, removing or disabling links to restrict navigation space to irrelevant nodes, iv) by augmenting links with additional information about the node behind the link, with some form of annotation, v) by dynamically generating new, non-authored links based on the user's interests and/or current context (i.e., location) in the system. Since a considerable amount of works have been published based on these adaptation techniques, they are further discussed in the next sub-sections.

For example, an adapting toolbar a) predicts the user's most likely task and b) changes the presentation and organization of UI functionality to support user with this task as shown in Figure 7.


*Figure 7 - Adapting toolbar example taken from Microsoft Word*

Within this research stream, social navigation provides excellent opportunities for tailoring navigation advice to individual users' tasks, knowledge or abilities. When looking at social navigation in the real world we observe interesting phenomena. When conducting direct social navigation (e.g. communication with another person to solve a navigational task) it is often the case that "advice-givers" tailors their navigational instructions to the "advice-seeker" subconsciously. Of course, this tailoring may not always be a benefit, but if we can match the right giver and seeker the likelihood of success increases.

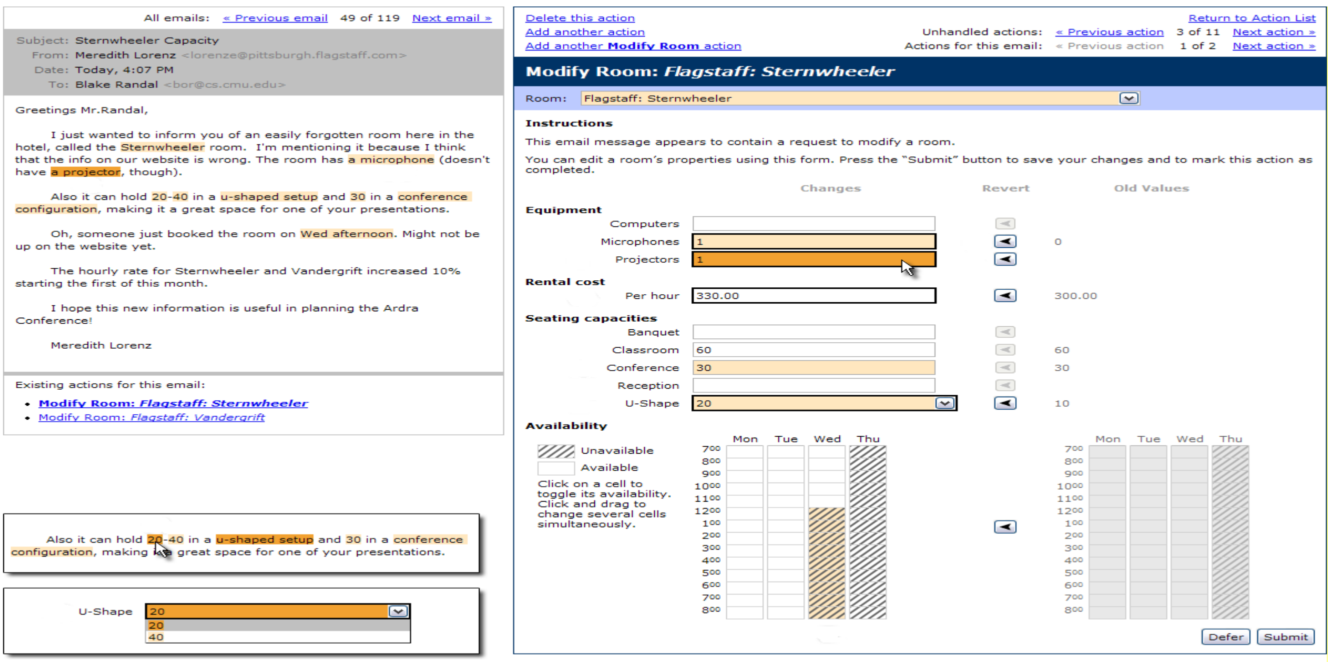


*Figure 8 - Assisted form-filling in RADAR*

For instance, the chat system PowWow7 uses something called on-line guides. These are expert PowWow users that have been granted “guide” status. Newcomers to the system can, at any time during the day go to a special “chat room” and ask guides questions concerning the system. This is an easy way to tailor (or personalize) the PowWow help system. (RADAR, n.d.) also support users to cope with email overload by a) identifying tasks requested in email messages b) classifying and prioritizing the tasks, and c) providing task-aware tools that partly automate task execution as seen in Figure 8.

#### 6.2.2.1 DIRECT GUIDANCE

This technique “guides” the user by suggesting the “next best” node to visit according to the user’s goals, preferences and knowledge. The suggested nodes are presented on the user’s interface by emphasizing existing hyperlinks or by generating a new “next” hyperlink which is connected to the suggested node. Direct guidance is popular in adaptive educational hypermedia systems where students get suggested nodes based on their level of knowledge on the specific subject. (Brusilovsky, 2003) reviewed several studies on direct guidance and demonstrated that users with poor knowledge on the domain can be best supported by direct guidance techniques. An interesting adaptive education hypermedia system that provides direct guidance is ELM-ART (Weber & Specht, 1997)

#### 6.2.2.2 LINK ORDERING

Adaptive link ordering prioritizes all hyperlinks of a node that are relevant to the user closest to the top. Despite its effectiveness in navigation times and steps reduction, an important drawback of adaptive link sorting is its limited applicability. Adaptive link sorting can only be used in hyperspaces where hyperlinks do not have a stable and predefined order. Thus, it can never be used with contextual links and rather difficult to be used for index pages or table of contents which usually have a predefined list of order.

In this respect, an appropriate context includes systems that contain non-contextual hyperlinks such as, adaptive news systems and commercial Web shops. Adaptive news systems typically recommend a prioritized list of news articles based on the modelled user’s interests and preferences. In the same way, commercial Web shops recommend a prioritized list of products based on the modelled user’s interests and product ratings. Link ordering is typically performed by content-based mechanisms.

#### 6.2.2.3 LINK HIDING

Link hiding aims to restrict navigation space by removing, hiding or disabling hyperlinks to irrelevant nodes. Link hiding has been very popular in the area of adaptive educational hypermedia systems that aim to protect the users from the complexity of the whole hyperspace and reduce their cognitive overload by hiding irrelevant to them nodes. For example, if the user has novice level of knowledge on a particular concept, the system restricts the user from navigating to it.

Variants of link hiding are: i) link hiding preserves the hyperlink’s functionality (i.e., navigate to the corresponding node), but removes all visual indications that it is a hyperlink (e.g., orange color and underlined), ii) link removal completely removes the hyperlink, and iii) link disabling removes the functionality of the hyperlink.

#### 6.2.2.4 LINK ANNOTATION

Link annotation augments the hyperlink with additional information about the node behind the annotated hyperlink, with some form of annotation. Link annotations are provided with different visual signs, for example different icons, different color and intensity of anchors, or different font sizes. Furthermore, Web technologies enabled adaptive Web systems to annotate hyperlinks with verbal annotations on hyperlink mouse-overs, for example display information on the browser's status bar or as a "balloon" over the hyperlink when the user moves the mouse pointer over the hyperlink.

#### 6.2.2.5 LINK GENERATION

Link generation has been very popular in adaptive Web systems, due to the rapid increase of open corpus document collections. Link generation dynamically creates new, non-authored hyperlinks on a webpage.

Link generation is popular in the field of adaptive navigation support systems and Web recommender systems for the dynamic generation of links that are useful within the current context to the current user. Web recommender systems attempt to recommend a prioritized list of relevant to the user items, typically based on the user's interests. In this respect, Web recommender systems focus in the underlying technology. On the other hand, adaptive navigation support systems focus on helping users to find their way through hyperspace by adapting links on a page. Link adaptation in adaptive navigation support systems take into account various features of the user, including user's interests, goals, knowledge, and current context (i.e., location in hyperspace). In all cases, navigation support techniques provide guidance that takes into account the user's current location in hyperspace (Brusilovsky & Millán, 2007). Thus, adaptive navigation support systems focus on the interface. Accordingly, although the difference between adaptive navigation support systems and Web recommender systems is not clear, an important difference between these two groups is that adaptive navigation support systems primarily focus on the user's current location in hyperspace and aims to guide the user by introducing additional hyperlinks that may be useful in the current context, while Web recommender systems primarily focus to recommend hyperlinks that are related with the user's short and long-term interests.

There also exists a small class of systems that generate hyperlinks based on user's interests and current location, for example (Amazon, n.d.) that recommend hyperlinks to products that were similarly rated or purchased by other users who viewed the current product.

## *6.3 Adaptive Visualization Techniques / Systems*

(Carenini, et al., 2014) investigated how the effectiveness of a visualization that is viewed by a user, can be increased with four different adaptive interventions (Boldind, De-Emphasizing, Reference Lines and Connected Arrows) and whether individual differences (locus of control, visual working memory, verbal working memory, perceptual speed), task complexity and intervention delivery time affect this effectiveness. The visualization used in the study for applying the different interventions was a bar graph, since bar graphs are ubiquitous, effective and performance has been proven to be affected by individual differences (Toker, et al., 2012). 62 participants took part in the study, where they had to answer multiple choice questions, by inspecting a bar graph. Experimental conditions varied, in task type (Retrieve Value RV and Compute Derived Value CDV), intervention type (4 interventions mentioned above or no intervention) and intervention delivery time (zero time or 500ms after the visualization was rendered). Participants also provided their intervention preference for both task types. Results regarding task performance indicate that all interventions helped participants solve tasks faster except the Reference Lines intervention. Another unique result that was not present in other studies, is that a link between verbal working memory and performance was detected, lower verbal working memory resulted in significantly less performance. For the more complicated CDV tasks, participants with higher values in visual working memory, verbal working memory and perceptual speed achieved better performance, while in simple RV tasks there was no significant difference. Moreover, when interventions were delivered dynamically there was a decrease in performance, possibly due to intrusiveness. In both delivery types all interventions had a better performance than no intervention, showing evidence that the interventions can help with performance. Participants also reported that all interventions were useful when compared to no intervention. Further on, participants with low or average visual working memory rated the usefulness of Average Reference Lines lower than users with higher visual working memory, this effect was explained as a visual destructor since this type of intervention poses a line on multiple bars. The paper concluded also that no single one intervention was superior than others, also there was no significant results linked to locus of control, likely because this personality trait is related to list-like visualizations with a containment metaphor (Ziemkiewicz, et al., 2011). An issue in this study is that no interaction effect was found between cognitive abilities and different interventions, the authors suggested that other interventions must be explored for users with low-medium cognitive measures.

Another adaptive system that uses a different approach to adaptation (behavior driven visualization recommendation) is the work of (Gotz & Wen, 2009). The paper proposes an algorithm that processes the visual exploration patterns of a user for detecting the user's current task and thus recommending a better visualization for the task at hand. The motivation of this study is that current recommender systems take into account the initial task / data of the user for recommending a visualization but do not take into account the evolving task of the user; this results in visual inertia (task evolves but users keep using the initial visualization). For understanding how users perform visual tasks, a number of participants performed realistic visual analytics tasks and the authors were able to analyze their interactions and extract patterns (Scan, Flip, Swap and Drill Down) which are made up of sequential analytic actions. Analytic actions were used to model the participants behavior, the vocabulary used was (inspect, filter and bookmark). Elicited patterns were a structure of analytic behavior and did not occur by chance, since 96% of users performed a pattern more than three times during the task. The detected patterns were then fed into the HARVEST web based visual analytics system for its underlying algorithm to be able to track users actions and detect analytical patterns they performed in real time. The system's recommendation algorithm could then use the detected patterns, the context and the current data set properties to infer a visualization that best fitted the current pattern and data. Once a new visualization recommendation was found the system indicated that a more reliable visualization was available, moreover, the user could choose whether to load the new visualization. The behavior driven visualization recommendation approach was evaluated for 2 patterns (Scan and Flip) on the HARVEST system using 20 participants that had to go through 6 tasks (half without recommendations disabled). Participants also evaluated the system and the recommendations on a Likert-scale. With recommendations enabled the users had faster completion times and less errors when compared to tasks with the recommendations disabled. Moreover, positive feedback was provided for the recommendations and system, a drawback of this system and the current evaluation is that it only accounted for two patterns and thus could not cope with more complicated ones.

Another approach to adapting a visualization is by user goals. (Brusilovsky, et al., 2006) presents a system that contains a spatial similarity-based visualization with adaptive icon annotations that allows students to locate examples relevant to their learning goal. The visualization arranged educational examples on a 2D map according to their similarity and the annotation icon on each example informed the student on the progress made with that example and whether an example was to be explored next. The motivation to the study, is how a user can access the right learning material given that dozens of examples are available the same time. The proposed visualization was zoomable and could display all available examples, moreover, the visualization was built with the spring model that arranged the examples on the visualization as nodes, positioning them according to its "*forces formula*" (similarity in this case). The similarity for each of the educational examples was calculated using the TF-IDF scheme for applying term weights for each example, and then the cosine similarity coefficient for measuring the similarity between the examples. A drawback to the above similarity measure is that the cosine similarity coefficient cannot account synonym words or account for context. Moreover, the next adaptive element in the visualization was the annotation icon on each example node that was adapted using a functionality called knowledge-based indexing (accounts on what the student already completed and prerequisites for each example). Each annotation icon had 2 states prerequisite-based (cannot access example) and progress-based (ready to be accessed / displays progress on example) for further informing the student.

Visualization adaptation not only helps with performing tasks faster or reaching a goal effectively, but in some contexts, adaptation also provides security as well. In the context of monitoring systems such as Security Information and Event Management (SIEM) (Yelizarov & Gamayunov, 2014) proposed a visualization of hosts that adapts according to the current cognitive load of the user for increasing efficiency when dealing with system threats (decreases extraneous load and increases effective load). The motivation to the study is that SIEM systems processes and display to the user thousands of transactions per second, without taking the operators continues monitoring and timely decision making into account, thus increasing the chance of costly errors. The proposed system was web based and fetched data from a SIEM server for building the adaptive network map of hosts and their events. The system initially collected the optimal psychological condition of an operator (perception speed and working memory) and then through the user's mouse and keyboard interaction it recalculated the operator's cognitive capabilities using an interaction interval and decides if the user was overloaded using the dual-task paradigm. Then for adaptation, the system calculated the importance of every element i.e. host (event severity, urgency) on the visualization and using the user's cognitive load value, it highlighted the most significant hosts and dimed the rest (important elements, more opacity). The system was evaluated with 8 IT operators that had to deal with a number of attacks and try to minimize the damage on the system through the proposed system, the tests were done with adaptation enabled and disabled. The adaptation resulted in an increased operator-interface efficiency of 42%, with the number of average damaged points decreasing from (25,873.75 disabled adaptation) to (18,215.12 enabled adaptation).

(Setlur & Stone, 2016) presented a different approach to visualization adaptation by using linguistic means for applying coloring to the bars (categories) of a bar graph, for minimizing the user's cognitive load when processing the visual information. The authors argue that if a visualization contains colorable terms i.e. Crayola colors, and the representative colors are incompatible, then there is a possibility of cognitive interference. The proposed algorithm uses the terms that define a given set of data to infer the potential colors. Since not all data categories can be colored the first step of the algorithm was to calculate the colorability of objects in a category before deciding on the color. For finding the colorability of a term the algorithm combined the term with the eleven basic colors e.g. milk black, milk white etc. and then passed this data into Google n-grams for finding the co-occurrence of term and color from a huge dataset of corpus aiming for a result of NPMI larger than or equal to 0.5. If the condition was met for a color / term combination, then the matched color/s were potential coloring candidates for the term. The next step of the algorithm was to determine the actual color. For determining the color, the algorithm took the results from Google n-grams and passed them as a query to the Google Images API with a number of parameter constraints (rank clipart images higher than photos, dominant color filter and confidence score larger than or equal to 0.65). After retrieving the images from the API clustering was performed for finding the dominant color. A limitation to the above algorithm is that it cannot distinguish between the context of a term i.e. apple (fruit or brand). The authors therefore adapted it to account for context as well by calculating the semantic relatedness of a provided phrase that describes the category to be colored, to the symbol synet in Wordnet (Lexical database for English) for finding the Least Common Subsumer, e.g. brands and companies are associated with logo. Then the highest scoring symbolic word was used to issue a query on Google Images targeting symbolic clipart. A drawback of the proposed algorithm is the fact that it cannot account for terms like feelings, since Google n-grams has limited coverage for abstract terms.

Finally, as we have seen in the analysis of modern DA / BI Platforms, research has also delved into algorithms that aids us in picking the right chart type in general. In order to evaluate how different users, with distinct usage patterns, cognitive traits and specific goals of analysing and using the data through visualizations, we need first to understand what types and interventions are available and how they can correlate the fore mentioned dimensions. A large number of different types of visualizations are utilized by data analysts depending on different parameters such as the nature of the data (e.g. geographical and partial data vs statistical data), the number of parameters, the medium through which it will be presented, if the data will be correlated, the end user, if the data is real time or historical or both etc. The main uses of a chart can be divided into general directions (goal-directed actions) like comparison, trend and distribution. Organizations have attempted to catalogue all the different types of visualizations and provide guidelines. Multiple available resources provide trees that help users pick the right chart for the goal they are trying to achieve. One of the largest online catalogues of visualizations can be found on the Data

Viz Project (datavisproject, n.d.) in Figure 9 or the Data Visualization Catalogue (dataviscataloque, n.d.) in  Figure 10.

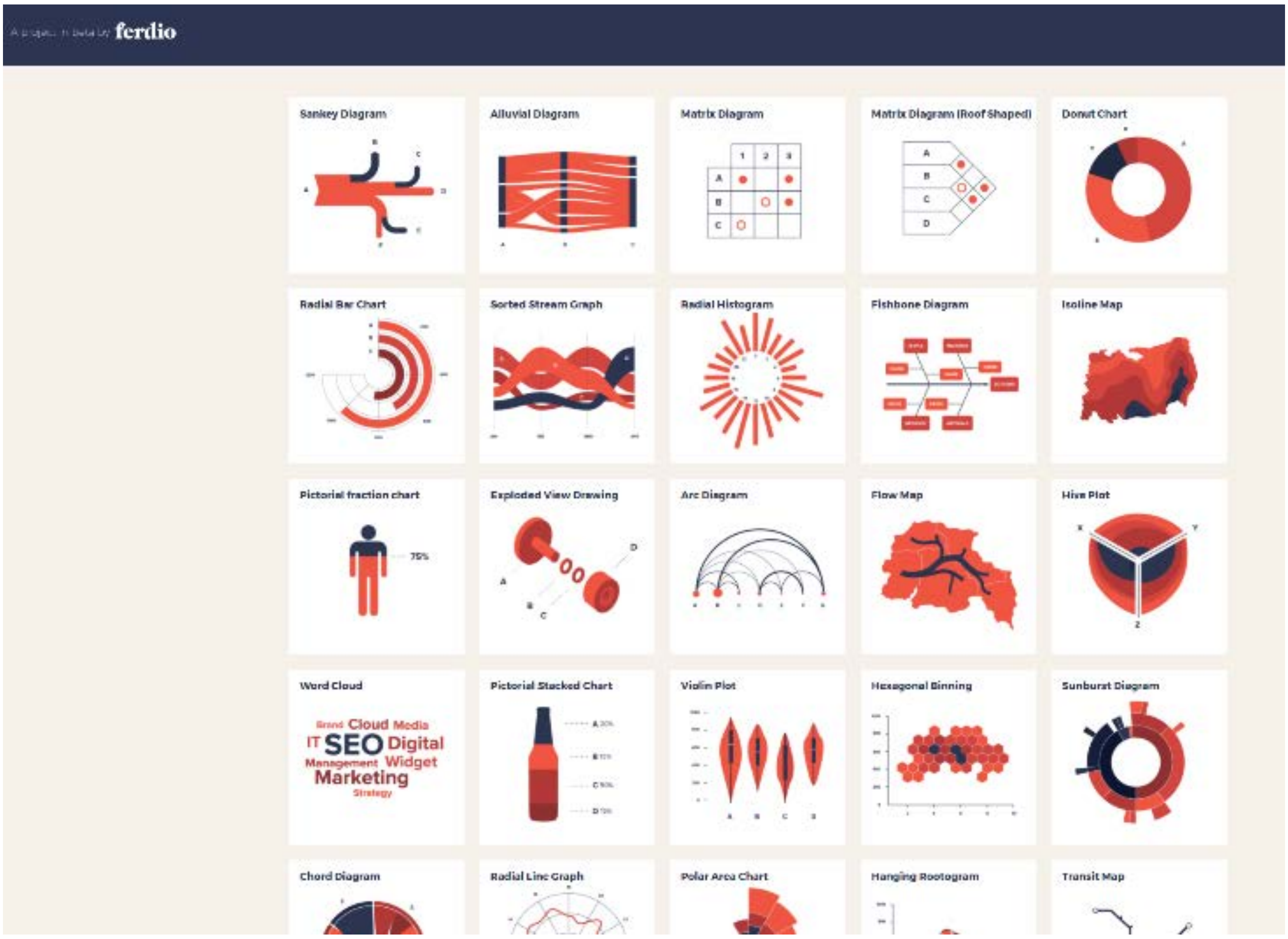


*Figure 9 - Snapshot from the Data Viz Project*


*Figure 10 - Snapshot from the Data Visualization Catalogue*

However, the main issue with these graphs is the unsupported claim and lack of justification behind their choices. Upon closer inspection, some research aims to back these assertions; (Doumont & Vandenbroeck, 2002) present a breakdown of each chart creation and comprising elements. While the creation purpose is offered, it does not necessarily mean that a certain chart is the most effective when used in its original creation context.

In summary, the table below (see Table 1) demonstrates the main uses for each chart type according to the goal-directed actions which might be related to a specific request for data analysis. A more detailed break-down can be found in Figure 11, Figure 12 and Figure 13.

*Table 1 - Chart types according to goal-directed actions for data analysis (exploration)*

| Chart Type | Comparison | Distribution | Composition | Trend | Relationship | Table |
|---|---|---|---|---|---|---|
| Alternating Rows Table | X | | X | | X | X |
| Bar Chart | X | X | X | | | |
| Bubble Chart | X | | | | X | |
| Bullet Bar Chart | X | | | | | |
| Circular Area Chart | X | | | | | |
| Column Chart | X | X | | X | | |
| Column Histogram | | X | | | | |
| Column Line Chart | | | | X | X | |
| Groupings Table | X | | X | | X | X |
| Line Chart | X | | | X | | |
| Line Histogram | | X | | | | |
| Pie Chart | | | X | | | |
| Pie Chart with Highlight | | | X | | | |
| Quartiles Table | X | | X | | X | X |
| Scatterplot Chart | | X | | | X | |
| Stacked Area Chart | | | X | | | |
| Stacked Bar Chart | X | | X | | | |
| Stacked Column Chart | | | X | X | | |
| Table | X | | | | | |
| Waterfall Chart | | | X | | | |

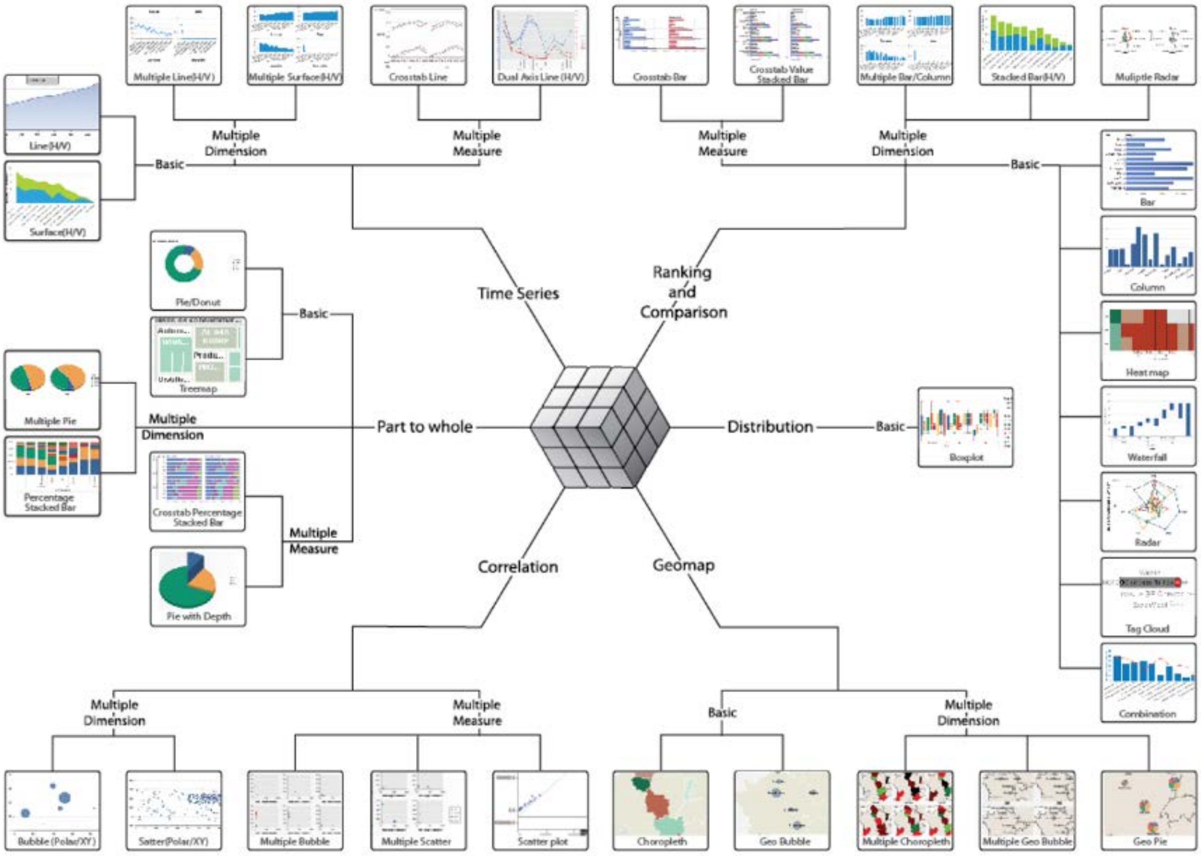


*Figure 11 - CVOM Charting Map*

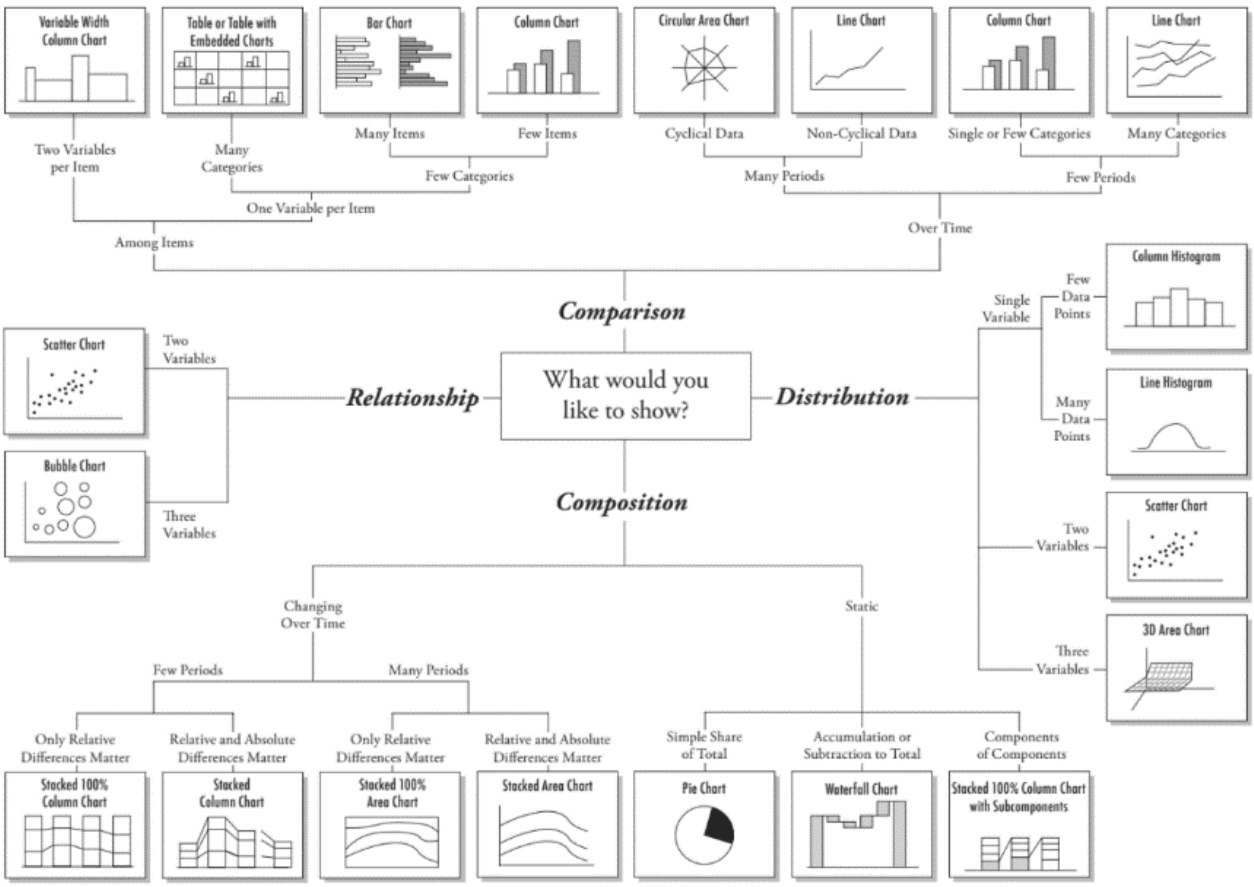


*Figure 12 - Chart Suggestions – A Thought Starter*

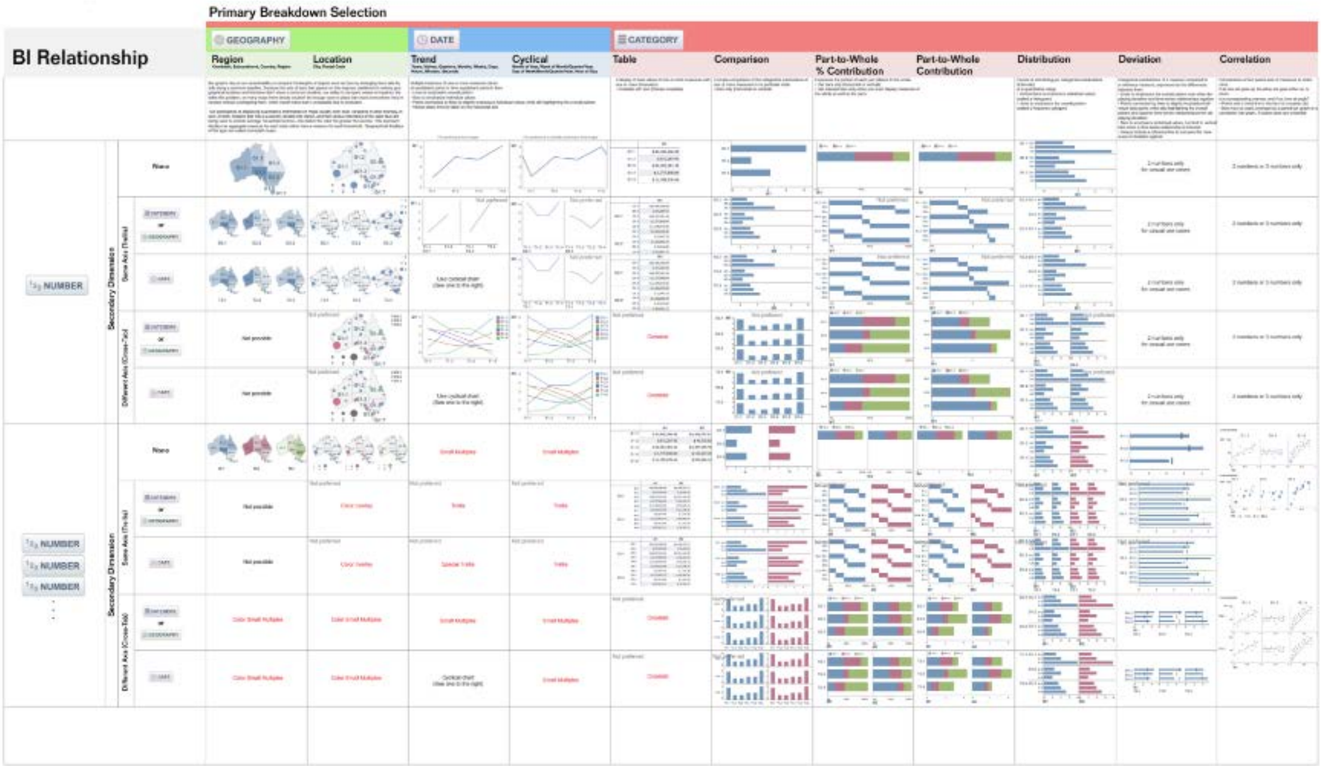


*Figure 13 - Visualization Suggestion Matrix Based on User Selection of Table Columns*

## 6.4 Visual Guidance

Visual guidance consists of techniques that aim to guide the user to better understand the underlying visual data connections and correlations, as this can be an overwhelming process if the user lacks some knowledge about visual analytics. A significant work on guidance in visual analytics is the work of (Ceneda, et al., 2017) which established a model that facilitates in depth reasoning about visual guidance methods that guide or assist users in visual analytics. The said model is built on top of the ( Wijk, 2006) model, a model that was used to understand whether a visualization method was worthwhile by inspecting factors like, the value of knowledge produced from it - “a great visualization method is one that is used by people, who use it routinely to obtain highly valuable knowledge, without having to spend time and money on hardware, software and effort”. It must also be noted that the ( Wijk, 2006) model illustrates a circular data exploration process, where the user expands his/her current knowledge by (1) viewing a visualization of a predefined specification, (2) receiving knowledge that is also affected by perception, human differences and already acquired knowledge, (3) interactively exploring the visualization and changing the current specification for further exploration (going back to 1 where the new specification selected updates the visualization). (Ceneda, et al., 2017) further defines guidance in visual analytics as a computer-assisted process that aims to actively resolve a knowledge gap encountered by users during an interactive visual analytics session. Moreover, (Ceneda, et al., 2017) expands ( Wijk, 2006) model, to a model of guided visual analytics by adding the Guidance Generation component that has inputs (the user’s knowledge, already explored visualizations, the original data, interaction history, domain conventions or models) and outputs (visual cues, options or alternatives, a new specification). The outputs of this function describe the provided guidance, that aims to maintain an environment in which users can progress effectively. Further on, the study, defined and explored the three main characteristics of guidance (knowledge gap, inputs and outputs as explained above in the function and how those outputs are to be conveyed to the user, guidance degree) and how those can be applied with the model in a real use case. One limitation to the proposed model though is that it does not take into consideration the user's mind map or his/her personal exploration techniques.

### *6.5 Indirect Collection of User Characteristics for Adaptation*

The study of (Conati, et al., 2011) has a long-term goal of adapting visualization systems to specific needs of the individual user. The main aim of the current study though is to investigate data sources that can capture user characteristics that affect visualization perception in real time for performing visualization adaptation, (1) through selecting a different visualization during the exploration process and (2) by providing adaptive help during exploration i.e. by drawing the attention of the user at a specific area. Moreover, authors suggest that the proficiency of a user with a visualization can be collected through interaction behaviors such as eye movement. An experiment took place where participants had to perform simple analytical tasks (comparisons, finding extreme values and computing derived values) using a bar graph and a radar graph while their completion time, answers and eye gaze data was collected. After the experiment, retrospective verbal protocols of user performance were collected by asking participants to verbalize how they reached their answers (used for coding verbal protocols of confusion and strategies that participants use to provide an answer). Finally, the authors proposed that using all the collected data (task correctness, completion time and coded verbal protocols) they can label eye gaze data and use them to build a classifier that can identify through eye gaze data, patterns of suboptimal visualization processing for calling visualization adaptations.

Cognitive styles are an important factor that can be used for tailoring interactive systems for specific users through adaptation and personalization. The study of (Raptis, et al., 2017) introduces a multifactorial model, for implicit elicitation of cognitive styles that minimizes the time-consuming task of explicit, in lab, paper based, non-real-time elicitation of human cognitive styles. We consider this work significant as it uses the principal stages of information processing (visual scanning and visual processing) through eye tracking and classification for inferring cognitive styles (Field Dependent Independent) specifically, while not being limited to a specific application domain. The proposed multifactorial model is made up of three factors, which are interconnected through eye tracking data (1) human cognition factor in the model is the cognitive style of FDI, (2) visual behavior factor is the visual differences between field independent and field dependent individuals, and (3) activity factor is how activity i.e. visual search tasks and visual decision making tasks affect the visual behavior of field independent and field dependent individuals. The goal of the study was to initially identify measurable visual behavior differences (number of fixations, gaze entropies, scan paths) among participants with different cognitive styles and then using classification, to infer a participant's cognitive style using the identified specific gaze measures. Two feasibility studies were performed, where participants were classified as field independent and field dependent using GEFT and had to perform two different visual activities (visual search and visual decision-making). Eye movement results were processed and transformed into specific eye-tracking measures. Based on gaze measures (gaze entropy) of the visual search activity a training learning model was formed on the Naïve Bayes classifier that yielded 81% accuracy when tested with a new dataset of gaze entropy measures. Moreover, based on gaze measures (quantity and duration of fixations on the first of the tasks) of the visual decision-making activity a training learning model  was formed on the Naïve Bayes and Logistic Regression classifiers that yielded 90% and 95% accuracy respectively when tested with a new dataset of gaze fixation count and duration measures. Finally, the study provided evidence that individual differences in cognitive styles are indeed reflected in a quantitative manner on eye gaze data, while users of a system perform various types of activities with varying characteristics.

A similar approach to the one of (Raptis, et al., 2017)  for implicitly capturing human differences through eye gaze data is the work of (Steichen, et al., 2013) with a difference, that instead this approach is specific to information visualization and tries to predict the user's visualization task and cognitive abilities (perceptual speed, verbal working memory and visual working memory) while a user is interacting with a visualization for informing adaptive visualization systems in real-time. In the experiment 35 participants completed cognitive tests for eliciting the abovementioned cognitive skills and then answered questions while extracting data from bar graphs and radar graphs. Questions varied in task type (retrieve value, filter, compute derived value, find extremum and sort) and complexity (single and double). Using a number of extracted gaze features, the authors used Linear Regression for training their models for their experiments. Moreover, the models were validated using 2 sets of features (1) including the areas of interest and (2) excluding the areas of interest. As a result, in all their predictions the set including the area of interest features helped in achieving more accurate predictions and thus concluding that, is a benefit to an adaptive visualization system to be aware of the currently loaded visualization and its areas of interest. The average accuracies of the experiments were not high enough to be used in a live system, but the authors argue that extra feedback (i.e. scan path patterns) to the system would have increased the accuracies. Other than that, important findings were detected in the correlations of gaze features and the target variables (task, complexity and cognitive styles). The classifier for predicting task was improved to an accuracy of 54% by combining two pairs of tasks together (derived value – filter and find extremum - sort) as they are similar in the way of solving. The task complexity classifier was the most accurate 85%, features that contributed to this include, the legend, since, as complexity increased the use of the legend also increased, this poses an important finding for adapting the legend when the task becomes more difficult. Classification results for cognitive abilities varied from 56%-60%  accuracies. Moreover, the peak accuracy for those experiments was found after 20%-40% of the data was observed meaning that cognitive abilities mostly affect a user's gaze patterns during the initiation of a task. Working memory was correlated with time to first fixation (label, text and high area of interest) – high visual working memory participants had lower time to first fixation. Verbal working memory was correlated with (label, text area of interest) – high verbal working memory participants spent less time in text area of interest. Finally, features that correlated with perceptual speed were the legend and the label area of interest – high perceptual speed participants had a lower number of fixations in the legend, a shortest longer fixation and a higher fixation rate.

In the context of recommender systems (Oard & Kim, 1998) proposes a methodology that allows for collecting user feedback implicitly, without for example requiring the user to provide explicit ratings about something on an ordinal or qualitative scale. The motivation to collecting feedback in an implicit manner is the removal of the generated cognitive load taken by a user who otherwise has to manually provide the feedback. The authors presented three categories of implicit feedback sources (Examination, Retention and Reference) and explained how each can be used by providing examples. Examination observes what things people select i.e. messages viewed, how much time people spend on a section of the page, moreover it captures their scrolling behavior through edit wear, can detect repetitions and purchases of objects or subscriptions which denote value. Then the Retention category observes behaviors that suggest that an item will be used in the future again by the user. Examples of retention are bookmarks (organized or default manner of saving matters). Examples of organized saving are saving an email in a custom folder. This provides that the user ascribes specific value to this object. Printing is another retention observation discussed. Moreover, the Retention category can also denote less value for an object, i.e. deletion of an email, assuming that the default behavior is to retain emails. The last category, Reference, explains links between two objects. Examples are when a mail is forwarded there is a link between the new message and the one being forwarded, also a link exists between hypertext links from one page to another, citations in papers, cutting a portion of one document into another etc. The authors finally presented two strategies of implementing implicit feedback into recommender systems. A key characteristic of those strategies is inference and prediction, the two strategies shown, structure the two terms in a different manner and the hybrid approach of using those strategies is suggested.

### *6.6 Personalized Visualizations*

(Mouine & Lapalme, 2012) proposes a system that provides its users with personalized weather visualizations that aims to remove the user's need of having to scan a mass of information. The proposed personalization method relies on predicting the user preferences and needs, using the history of the preferences of similar users. Interactivity is an important part of the proposed system, as user visualization interactions i.e. filtering, type of selected visualization, level of selected detail are used to train the system for predicting user preferences and generating custom visualizations. Using K-Means clustering the system is able to create clusters of user groups based on their similarity of interaction preferences and other user profile data like (location based on IP, selected language, time of connection depending on location and the season). Interaction preferences captured by the system are the final settings selected by the user. Moreover, the proposed system categorizes user based on their device before clustering is applied since visualizations vary depending on the screen of the device. Any new visualization requested by the user would then be influenced by the visualization preferences of other users in the same group, influence degree is affected by the variable d which is the distance of a user from the other elements of the cluster. The final report generated by the system is proposed to be built using a method called document planning, more specifically the visualization is said to be generated using multiple parameters such as visualization preferences inferred from clustering, user choices or device type. Lastly is mentioned that the proposed system would also highlight warnings about bad weather, but this is not related to the personalization aspect of the system.

### *6.7 Other Adapted / Personalized Systems*

Given the multidimensional character of adaptation and personalization research and paradigms, building a complete adaptive system is a challenging endeavour. Thus, the literature reveals a high number of research works that focus and investigate targeted issues than complete personalization systems. For example, incorporating human factors in the design of personalized user authentication mechanisms requires first investigating whether specific human factors affect user interactions in authentication-related tasks. In this context, this section presents a selection of adaptation and personalization systems and architectures starting from recent systems to early and pioneering works. Main aim is to acquire the knowledge on challenges, difficulties, techniques and best practices of other domains so to build upon and adopt these lessons learned more comprehensively to the requirements and constraints of IDEALVis.

#### 6.7.1 MOOCLET

(Williams & Heffernan, 2015) demonstrated the MOOClet formalism that allows for discovering how to adapt or personalize various technology components like emails and web components by performing randomized experiments with variables associated with a user model. This study demonstrated how this formalism was used to personalize an email for increasing the rate of replies. The approach taken with the formalism was to (1) run a randomized comparison of the different versions of an email (micro-designs); (2) evaluate each version over another taking in consideration the (known) user characteristics for finding which versions are beneficial to which subgroup of users; (3) dynamically change the so called policy of which email is sent to which user group. Adaptations made by the formalization (change the email to be sent to a group of users) are based on analysis of how different users with different characteristics act (responded to email) upon each micro-design (email). The main MOOClet formalism architectural components are MOOClets (a number of micro-designs), User Variable Store that contains the known user characteristics, Experimental Policy that dictates how different micro designs are delivered to users and Adaptive Personalization Policy that acts with data collected for matching which user characteristics match which micro-design. In the current study, an experiment was conducted with 27 different emails as micro-designs. The mails were distributed (in intervals) to 5500 students of which the characteristics were known. Data about who responded was collected at each interval and the conditions in the Adaptive Personalization Policy were modified. The results indicated that the response rate was increased by 50%. While this is a generalizable method for adapting multiple types of systems, it requires a large amount of users until “good-enough” personalization rules are formed, something that cannot be tolerated by systems that demand the adaptation to start early in the process without pre-existing experimental data being available.

#### 6.7.2 PAC

PAC (Personalized Authentication and CAPTCHA) (Belk, et al., 2015) is an extensible personalization framework that adapts and personalizes specific design factors of user authentication and CAPTCHA mechanisms based on a set of human cognitive factors. In particular, the personalization framework follows a two-phase method for adapting and personalizing the user authentication and CAPTCHA task as follows: i) adapt the type of the security mechanism (textual or graphical) based on users' cognitive styles (i.e., Verbal/Imager and Wholist/Analyst); and ii) adapt the complexity level of the security mechanism (number of characters/images) based on users' cognitive processing abilities (i.e., limited/enhanced).

#### 6.7.3 PERSONA-WEB

PersonaWeb (Germanakos, et al., 2015) focuses on adapting and personalizing content and functionality of E-Commerce environments based on human cognitive factors. In the frame of the PersonaWeb system, new adaptation effects have been proposed for adapting the visual and interaction design of E-Commerce product views. An additional sub-system, called PersonaCheck (Constantinides, et al., 2015) has been included that is responsible to recommend the "best-fit" checkout process design based on the way individuals process and mentally organize information (holistically or analytically). PersonaWeb experimental studies have shown that users' task completion efficiency and effectiveness improve when E-Commerce product views and checkout designs are adapted to the users' cognitive characteristics, in contrast to the original, baseline design.

### 6.7.4 ADAPTIVE NOTIFICATIONS IN VIRTUAL COMMUNITIES

In the work of (Kleanthous & Dimitrova, 2012) a framework has been proposed for supporting knowledge sharing in virtual communities through adaptive notifications. It employs a novel computational approach for community-tailored support underpinned by the area of organizational psychology, aiming to facilitate the functioning of the community as a whole entity. The framework makes use of a community model that represents the community based on key processes (i.e., transactive memory, shared mental models and cognitive centrality) aiming to derive knowledge sharing patterns from community log data that are used to generate adaptive notifications.

### 6.7.5 EKPAIDEION

EKPAIDEION (Tsianos, et al., 2008) is an adaptive educational hypermedia system that adapts and personalizes the content presentation and navigation support within computer-based educational environments. The system utilizes a human factor based user model that incorporates a combination of human cognitive factors based on a novel, unified theoretical model. The theoretical model entails a set of elementary cognitive processes (visual attention, speed and control of processing, working memory), cognitive styles and emotional factors (anxiety, emotional regulation) and accordingly adapts and personalizes the content presentation, learners' support, navigation menus as well as provides adaptive navigational support during user interactions in E-Learning environments.

### 6.7.6 ADAPTIVEWEB

The AdaptiveWeb system ( Germanakos, et al., 2008) was one of the early systems of the authors that aimed to personalize content and functionality of interactive systems based on intrinsic human factors. In particular, AdaptiveWeb is a Web-based adaptation and personalization system that is based on a comprehensive user model, incorporating "traditional" user characteristics (i.e., name, age, education, experience, profession, etc.) and intrinsic human factors such as the users' perceptual preference characteristics (visual, cognitive and emotional processing parameters). According to the user model, the system provides adaptive content presentation and adaptive navigation support in the context of an E-Learning environment aiming to assist users during information processing, comprehension and assimilation.

### 6.7.7 M-PERSONA

mPERSONA (Panayiotou & Samaras, 2004) is a flexible personalization system for the wireless user that takes into consideration user mobility, the local environment and the user and device profile. The system utilizes the various characteristics of mobile agents to support flexibility, scalability, modularity and user mobility. It avoids tying up to specific wireless protocols (e.g., WAP) by using, as much as possible, autonomous and independent components. To achieve a high degree of independence and autonomy mPERSONA is based on mobile agents and mobile computing models such as the “client intercept model”.

#### 6.7.8 INSPIRE

INSPIRE (Papanikolaou, et al., 2003) is an Adaptive Educational Hypermedia system, which emphasizes the fact that learners perceive and process information in very different ways, and integrates ideas from theories of instructional design and learning styles. Its aim is to make a shift towards a more “learning-focused” paradigm of instruction by providing a sequence of authentic and meaningful tasks that matches learners’ preferred way of studying. INSPIRE, throughout its interaction with the learner, dynamically generates learner-tailored lessons that gradually lead to the accomplishment of learner’s learning goals. It supports several levels of adaptation: from full system-control to full learner-control and offers learners the option to decide on the level of adaptation of the system by intervening in different stages of the lesson generation process and formulating the lesson contents and presentation. Both the adaptive and adaptable behaviour of INSPIRE are guided by the learner model which provides information about the learner, such as knowledge level on the domain concepts and learning style. The learner model is exploited in multiple ways: curriculum sequencing, adaptive navigation support, adaptive presentation, and supports system’s adaptable behaviour.

#### 6.7.9 SQL-TUTOR

SQL-Tutor (Mitrovic & Martin, 2002) is a knowledge-based teaching system which supports students learning SQL. The intention was to provide an easy-to-use system that will adapt to the needs and learning abilities of individual students. The tailoring of instruction is done in two ways: by adapting the level of complexity of problems and by generating informative feedback messages.

#### 6.7.10 PROTEUS

Proteus (Anderson, et al., 2001) is a system that constructs user models using artificial intelligence techniques and adapts the content of a website taking into consideration also characteristics of the wireless connection. The Proteus Web-site personalizer performs a search through the space of possible websites. The initial state is the original website of non-adapted pages. The state is transformed by any of a number of adaptation functions, which can create pages, remove pages, add links between pages, etc. The value of the current state (i.e., the value of the website) is measured as the expected utility of the website for the current visitor. The search continues either until no better state can be found, or until computational resources (e.g., time) expire.

#### 6.7.11 WBI - WEB BROWSER INTELLIGENCE

Web Browser Intelligence (WBI, pronounced “WEB-ee”) (Maglio & Barrett, 2000) is an implemented system that provides a loosely confederated group of agents on a user's workstation capable of observing user actions, proactively offering assistance, modifying resulting web documents, and performing new functions. For example, WBI will annotate hyperlinks with network speed information, record pages viewed for later access, and provide shortcut links for common paths. WBI is an architecture in which small programs, or agents, connect to the information stream by registering their trigger conditions and then performing operations on the stream. This structure provides rich opportunities for personalizing the web experience by joining together personal and global information, as well as enabling collaboration among web users.

#### 6.7.12 ARCHIMIDES

ARCHIMIDES (Bogonikolos, et al., 1999) personalized the search results of users according to their interests. The system was based on agent technologies aiming to provide adaptive and personalized navigation to users within Web-based environments. Given a set of keywords that characterize the content on a Web server, ARCHIMIDES retrieves information intelligently and then constructs a personalized version in the form of an index pointing to pages that present some interest to the user.

#### 6.7.13 TANGOW

TANGOW (Carro, et al., 1999) is a tool for developing Internet-based courses, accessible through any standard WWW browser. Courses are structured by means of Teaching Tasks and Rules which are stored in a database and are the basis of TANGOW guidance ability. In TANGOW a Student Process is launched for each student connected to the system. Each Student Process consists of two main modules: a Task Manager that guides the students in their learning process, and a Page Generator that generates the HTML pages presented to the student. The Student Process also maintains information about the actions performed by the student when interacting with the course in the Dynamic Workspace. This information is used by TANGOW to adapt the course contents to the student's learning progress. TANGOW has also information about student profiles, which is used to select, at run-time, the contents of each HTML page presented.

#### 6.7.14 INTERBOOK

InterBook (Brusilovski, et al., 1998) is a tool for authoring and delivering adaptive electronic textbooks on the World Wide Web. InterBook provides a technology for developing electronic textbooks from a plain text to a specially annotated HTML. InterBook also provides an HTTP server for adaptive delivery of these electronic textbooks over WWW. For each registered user, an InterBook server maintains an individual model of user's knowledge and applies this model to provide adaptive guidance, adaptive navigation support, and adaptive help.

#### 6.7.15 AHA!

AHA (Bra & Calvi, 1998) is an open Adaptive Hypermedia Architecture that is suitable for many different applications. This system maintains the user model and filters content pages and link structures accordingly. The engine offers adaptive content through conditional inclusion of fragments. Its adaptive linking can be configured to be either link annotation or link hiding. Even link disabling can be achieved through a combination of content and link adaptation.

#### 6.7.16 SKILL

SKILL (Neumann & Zirvas, 1998) is a scalable Internet-based teaching and learning system. The primary objective of SKILL is to cope with the different knowledge levels and learning preferences of the students, providing them with a collaborative and adaptive learning environment utilizing new World Wide Web technologies. Basic components of SKILL are course material based on concepts organized in an ordinal rating derived from pre-requirements, an annotation facility suited for collaboration work, and a configuration environment for tailoring the system. Topics discussed include: (1) SKILL functionality, including adaptivity/progress control and collaboration through annotations and course extensions; (2) components, including security, document management, and tutoring components; (3) implementation issues; and (4) related work.

#### 6.7.17 ELM-ART II

ELM-ART II (Weber & Specht, 1997) is an intelligent interactive textbook to support learning programming in LISP. ELM-ART II demonstrates how interactivity and adaptivity can be implemented in WWW-based tutoring systems. The knowledge-based component of the system uses a combination of an overlay model and an episodic user model. It also supports adaptive navigation as individualized diagnosis and help on problem solving tasks. Adaptive navigation support is achieved by annotating links. Additionally, the system selects the next best step in the curriculum on demand. Results of an empirical study show different effects of these techniques on different types of users during the first lessons of the programming course.

#### 6.7.18 BASAR

BASAR (Building Agents Supporting Adaptive Retrieval) (Thomas & Fischer, 1997) provides users with assistance when managing their personal information spaces. This assistance is user-specific and done by software agents called Web assistants and active views. Users delegate tasks to Web assistants that perform actions on their views of the World Wide Web and on the history of all user actions.

# 7 Bibliography

Schrage, M., 2016. *Harvard Business Review.* [Online]

Available at: https://hbr.org/2016/08/how-the-big-data-explosion-has-changed-decision-making

[Accessed 05 01 2020].

IBM, n.d. *IBM, Bringing big data to the enterprise.* [Online]

Available at: https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html

[Accessed 02 01 2020].

Domo, n.d. *DOMO - "Data Never Sleeps 2.0".* [Online]

Available at: https://www.domo.com/learn/data-never-sleeps-2

[Accessed 05 01 2020].

Forrester, 2018. *Itelligencegroup - Forrester Research "Becoming an Insights-driven Company".* [Online]

Available at: https://us.itelligencegroup.com/hubfs/IB-NN-Analytics/Infographic-Forrester-Insights-Driven-Business.pdf?hsCtaTracking=19708524-effe-40f1-b888-

[Accessed 06 01 2020].

Sallam, R. et al., 2018. *Other Vendors to Consider for Modern Analytics and BI,* s.l.: Gartner.

Laney, D., Gassman, B., Sallam, R. & Yuen, D., 2013. *Predicts 2013: Business Intelligence and Analytics Need to Scale Up to Support Explosive Growth in Data Sources,* s.l.: Gartner.

Richardson, J. et al., 2020. *Gartner.* [Online]
Available at: https://www.gartner.com/doc/reprints?id=1-3TXXSLV&ct=170221&st=sb&ocid=mkto_eml_EM597235A1LA1&mkt_tok=eyJpIjoiTm1GbE5ESXhOemxsTUdSaSIsInQiOiJqOXU1amhJTzZheDVzWXZGV2hHcmx0WmwxSzJUUHdvZEZVc0hhRXowYWhJdGhQTkpxTVV5VG5Fc21xMG92SEpib044dmZVRjdLM3JaWVNtSVNxUEFPS
[Accessed 04 April 2020].

Liu, S. et al., 2017. Visualizing High-Dimensional Data: Advances in the Past Decade. *IEEE Transactions on Visualization and Computer Graphics ,* 23(3), pp. 1249 - 1268.

Cui , W., 2019. Visual Analytics: A Comprehensive Overview. *IEEE Access,* Volume 7, pp. 81555 - 81573.

TIBCO, n.d. *TIBCO.* [Online]
Available at: https://www.tibco.com/
[Accessed 19 05 2020].

PowerBI, n.d. *Power BI.* [Online]
Available at: https://powerbi.microsoft.com/en-us/
[Accessed 19 05 2020].

MicroStrategy, n.d. *MicroStrategy.* [Online]
Available at: https://www.microstrategy.com/us/product/analytics
[Accessed 19 05 2020].

ThoughtSpot, n.d. *ThoughtSpot.* [Online]
Available at: https://www.thoughtspot.com/
[Accessed 19 05 2020].

Sisense, n.d. *Sisense.* [Online]
Available at: https://www.sisense.com/
[Accessed 19 05 2020].

Looker, n.d. *Looker.* [Online]
Available at: https://looker.com/
[Accessed 19 05 2020].

OracleAnalyticsCloud, n.d. *Oracle Analytics Cloud.* [Online]

Available at: https://www.oracle.com/middleware/technologies/oracle-analytics-cloud.html

[Accessed 19 05 2020].

EinsteinAnalytics, n.d. *EinsteinAnalytics.* [Online]

Available at: https://www.salesforce.com/eu/products/einstein-analytics/overview/

[Accessed 19 05 2020].

YellowFin, n.d. *YellowFin.* [Online]

Available at: https://www.yellowfinbi.com/

[Accessed 19 05 2020].

Tableau, n.d. *Tableau.* [Online]

Available at: https://www.tableau.com/

[Accessed 19 05 2020].

Shneiderman, B., 1996. *The eyes have it: a task by data type taxonomy for information visualizations.* Boulder, IEEE.

Kerren, A. & Schreiber, F., 2012. *Toward the role of interaction in visual analytics Share on.* Berlin, Winter Simulation Conference.

Liu, S., Cui, W., Wu, Y. & Liu, M., 2014. A survey on information visualization: recent advances and challenges. *International Journal of Computer Graphics,* 30(12), p. 1373–1393.

Koć-Januchta, M. et al., 2017. Visualizers versus verbalizers: Effects of cognitive style on learning with texts and pictures – An eye-tracking study. *Computers in Human Behavior,* Issue 68, pp. 170-179.

Riding, R. & Douglas, G., 1993. The effect of cognitive style and mode of presentation on learning performance. *Educational Psychology,* 63(2), pp. 297-307.

Tsianos, N. et al., 2009. *Eye-Tracking Users' Behavior in Relation to Cognitive Style within an E-learning Environment.* Riga, Latvia, Ninth IEEE International Conference on Advanced Learning Technologies.

Mawad, F. et al., 2015. Influence of cognitive style on information processing and selection of yogurt labels: Insights from an eye-tracking study. *Food Research International,* Volume 74, pp. 1-9.

Toker, D., Conati, C., Carenini, G. & Haraty, M., 2012. *Towards Adaptive Information Visualization: On the Influence of User Characteristics.* Berlin, International Conference on User Modeling, Adaptation, and Personalization.

Lallé, S., Conati, C. & Carenini, G., 2017. *Impact of Individual Differences on User Experience with a Real-World Visualization Interface for Public Engagement.* New York, Conference on User Modeling, Adaptation and Personalization.

Toker, D., Conati, C., Steichen, B. & Carenini, G., 2013. *Individual user characteristics and information visualization: connecting the dots through eye tracking.* New York, Association for Computing Machinery.

Green, T. M. & Fisher, B., 2010. *Towards the Personal Equation of Interaction: The impact of personality factors on visual analytics interface interaction.* Salt Lake City, IEEE Symposium on Visual Analytics Science and Technology.

Ziemkiewicz, C. et al., 2011. *How locus of control influences compatibility with visualization style.* Providence, RI, USA, IEEE Conference on Visual Analytics Science and Technology (VAST).

Lee, S. et al., 2016. How do People Make Sense of Unfamiliar Visualizations?: A Grounded Model of Novice's Information Visualization Sensemaking. *IEEE Transactions on Visualization and Computer Graphics,* 22(1), pp. 499 - 508.

Allport, G. W., 1937. *Personality: A Psychological Interpretation.* 1 ed. s.l.:Henry Holt Later Printing.

Messick, S., 1984. The nature of cognitive styles: Problems and promise in educational practice. *Educational Psychologist,* 19(2), pp. 59-74.

Sternberg, R. J. & Grigorenko, E. L., 1997. Are cognitive styles still in style?. *American Psychologist,* 52(7), pp. 700-712.

Riding, R. & Cheema, I., 1991. Cognitive Styles—an overview and integration. *Educational Psychology,* 11(3-4), pp. 193-215.

Zhang, L.-f., 2004. Field-dependence/independence: cognitive style or perceptual ability?––validating against thinking styles and academic achievement. *Personality and Individual Differences,* 37(6), pp. 1295-1311.

Guisande, M. A., Páramo, M. F., Vacas, C. T. & S. Almeida, L., 2007. Field dependence-independence (FDI) cognitive style:An analysis of attentional functioning. *Psicothema,* 19(4), pp. 572-577.

Ziemkiewicz, C. & Kosara, R., 2008. The Shaping of Information by Visual Metaphors. *IEEE Transactions on Visualization and Computer Graphics,* 14(6), pp. 1269 - 1276.

Cleveland, W. S. & McGill, . R., 1984. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American Statistical Association,* 79(387), pp. 531-554.

Cleveland, W. S. & McGill, R., 1985. Graphical Perception and Graphical Methods for Analyzing Scientific Data. *American Association for the Advancement of Science,* 229(4716), pp. 828-833.

Nowell, L., Schulman, R. & Hix, D., 2002. *Graphical encoding for information visualization: an empirical study.* Boston, IEEE.

Hagh-Shenas, H. & Interrante, V., 2005. *Compositing color with texture for multi-variate visualization.* Dunedin, Proceedings of the 3rd International Conference on Computer Graphics and Interactive Techniques.

Cahill, M.-C. & Carter, Jr., R. C., 1976. Color Code Size for Searching Displays of Different Density. *Human Factors: The Journal of the Human Factors and Ergonomics Society,* 18(3), pp. 273-280.

Amar, R. A., Eagan, J. R. & Stasko, J., 2005. *Low-Level Components of Analytic Activity in Information Visualization.* s.l., IEEE Computer Society.

Dimara, E., Bezerianos, A. & Dragicevic, P., 2016. The Attraction Effect in Information Visualization. *IEEE Transactions on Visualization and Computer Graphics,* 23(1), pp. 471 - 480.

Stone, M. & Bartram, L., 2009. Alpha, contrast and the perception of visual metadata..

Carenini, G. et al., 2014. *Highlighting interventions and user differences: informing adaptive information visualization support.* New York, Association for Computing Machinery.

Gotz, D. H. & Wen, Z., 2009. *Behavior-driven visualization recommendation.* New York, Association for Computing Machinery.

Brusilovsky, P., Ahn, J.-w., Dumitriu, T. & Yudelson, M., 2006. *Adaptive Knowledge-Based Visualization for Accessing Educational Examples.* London, IEEE.

Yelizarov, A. & Gamayunov, D. Y., 2014. *Adaptive Visualization Interface That Manages User's Cognitive Load Based on Interaction Characteristics.* New York, Association for Computing Machinery.

Setlur, V. & Stone, M. . C., 2016. A linguistic approach to categorical color assignment for data visualization. *IEEE Transactions on Visualization and Computer Graphics,* 22(1), pp. 698 - 707.

Mouine, M. & Lapalme, G., 2012. *Using Clustering to Personalize Visualization.* Montpellier, IEEE Computer Society.

Ceneda, D. et al., 2017. Characterizing Guidance in Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics,* 23(1), pp. 111 - 120.

Wijk, . J. v., 2006. Views on Visualization. *IEEE Transactions on Visualization and Computer Graphics,* 12(4), pp. 421 - 432.

Conati, C. et al., 2011. *User-Adaptive Visualizations: Can Gaze Data Tell Us When a User Needs Them?.* San Francisco, Scalable Integration of Analytics and Visualization.

Raptis, G. E. et al., 2017. *Using Eye Gaze Data and Visual Activities to Infer Human Cognitive Styles: Method and Feasibility Studies.* New York, Association for Computing Machinery.

Steichen, B., Carenini, G. & Conati, C., 2013. *User-adaptive information visualization: using eye gaze data to infer visualization tasks and user cognitive abilities.* New York, Association for Computing Machinery.

Oard, D. W. & Kim, J., 1998. Implicit feedback for recommender systems.. *Proceedings of the AAAI workshop on recommender systems,* Volume 83.

Williams, J. J. & Heffernan, N. T., 2015. A Methodology for Discovering How to Adaptively Personalize to Users Using Experimental Comparisons. *SSRN Electronic Journal.*

Kosslyn, S. M., 1985. Review: Graphics and Human Information Processing: A Review of Five Books. *Journal of the American Statistical Association,* 80(391), pp. 499-512.

Tufte, E. R., 2001. *The Visual Display of Quantitative Information.* 2 ed. Connecticut: Graphics Press.

SMITH, S. L., 1979. Letter Size and Legibility. *The Journal of the Human Factors and Ergonomics Society,* 21(6), p. 661–670.

Mattturck, 2016. *Matt Turck.* [Online]
Available at: https://mattturck.com/big-data-landscape/#more-917
[Accessed 13 05 2020].

Columbus, L., 2016. *Forbes.* [Online]
Available at: https://www.forbes.com/sites/louiscolumbus/2016/08/20/roundup-of-analytics-big-data-bi-forecasts-and-market-estimates-2016/#21efeb026f21
[Accessed 10 05 2020].

Gentile, B., 2014. *Data Informed.* [Online]
Available at: http://data-informed.com/top-5-business-benefits-using-data-visualization/
[Accessed 05 08 2018].

D3JS, n.d. *D3.* [Online]
Available at: https://d3js.org/
[Accessed 10 05 2020].

Highcharts, n.d. *Highcharts.* [Online]
Available at: https://www.highcharts.com/
[Accessed 10 05 2020].

Yen, G. & Acay, D., 2009. Adaptive user interfaces in complex supervisory tasks. *ISA Transactions,* 48(2), pp. 196-205.

Horvitz, E. J., 1999. *Principles of mixed-initiative user interfaces.* Pittsburgh, Association for Computing Machinery.

Kühme, T., 1993. *A user-centered approach to adaptive interfaces.* New York, Association for Computing Machinery.

Keeble, R. & Macredie, R., 2000. Assistant agents for the world wide web intelligent interface design challenges. *Interacting with Computers,* 12(4), pp. 357 - 381.

Oppermann, R., 1994. Adaptively supported adaptability. *International Journal of Human-Computer Studies,* 40(3), pp. 455-472.

Wang, J. et al., 2010. Personalization of tagging systems. *Information Processing and Management,* Volume 46, pp. 58-70.

Smyth, B., 2007. *Case-Based Recommendation. In: Brusilovsky P., Kobsa A., Nejdl W. (eds) The Adaptive Web. Lecture Notes in Computer Science, vol 4321. Springer, Berlin, Heidelberg.* s.l.:s.n.

Chesnais, P., Mucklo, M. & Sheena, J., 1995. *The Fishwrap personalized news system.* Princeton, IEEE.

Yahoo, n.d. *Yahoo.* [Online]
Available at: https://www.yahoo.com/
[Accessed 10 05 2020].

MSN, n.d. *MSN.* [Online]
Available at: https://www.msn.com/
[Accessed 10 05 2020].

Googlenews, n.d. *Googlenews.* [Online]
Available at: https://news.google.com/
[Accessed 05 10 2020].

BRIDGE, D., GÖKER, M. H., McGINTY, L. & SMYTH, B., 2006. Case-based recommender systems. *The Knowledge Engineering Review,* 20(1), pp. 315-320.

Pazzani, M. J., 1999. A Framework for Collaborative, Content-Based and Demographic Filtering. *Artificial Intelligence Review,* 13(5-6), pp. 393-408.

Das, A. S., Datar, M., Garg, A. & Rajaram, S., 2007. *Google news personalization: scalable online collaborative filtering.* New York, Association for Computing Machinery.

Aggarwal, C. C. & Yu, P. S., 2002. *An Automated System for Web Portal Personalization.* Hong Kong, s.n.

Kamba, T., Sakagami, H. & Koseki, Y., 1997. ANATAGONOMY. *International Journal of Human-Computer Studies,* 46(6), pp. 789-803.

Sarwar, B. M., Karypis, G., Konstan, J. A. & Riedl, J., 2001. *Item-based collaborative filtering recommendation algorithms.* New York, Association for Computing Machinery.

Nores, M. L. et al., 2011. *Property-based collaborative filtering for health-aware recommender systems.* Las Vegas, IEEE.

Blanco-Fernández, Y., López-Nores, M., Pazos-Arias, J. J. & García-Duque, J., 2011. An improvement for semantics-based recommender systems grounded on attaching temporal information to ontologies and user profiles. *Engineering Applications of Artificial Intelligence,* 24(8), pp. 1385-1397.

Bobadilla, J., Ortega, F., Hernando, A. & Gutiérrez, A., 2013. Recommender systems survey. *Knowledge-Based Systems,* Volume 46, pp. 109-132.

Brusilovsky , P., 2001. Adaptive Hypermedia - User Modeling and User-Adapted Interaction. Volume 11, pp. 87-110.

Germanakos, P. et al., 2008. Realizing Comprehensive User Profile as the Core Element of Adaptive and Personalized Communication Environments and Systems. *The Computer Journal,* 52(7), p. 749–770.

Park, J. & Han, S. H., 2011. Complementary menus: Combining adaptable and adaptive approaches for menu interface. *International Journal of Industrial Ergonomics,* 41(3), pp. 305-316.

Kardaras, D. K., Karakostas, B. & Mamakou, X. J., 2013. Content presentation personalisation and media adaptation in tourism web sites using Fuzzy Delphi Method and Fuzzy Cognitive Maps. *Expert Systems with Applications,* 40(6), pp. 2331-2342.

RADAR, n.d. *RADAR.* [Online]
Available at: http://www.radar.cs.cmu.edu
[Accessed 10 05 2020].

Amazon, n.d. *Amazon.* [Online]
Available at: https://amazon.com
[Accessed 10 05 2020].

Brusilovsky, P., 2003. Adaptive Navigation Support in Educational Hypermedia: the Role of Student Knowledge Level and the Case for Meta-Adaptation. *British Journal of Educational Technology,* 34(4), pp. 487-497.

Weber, G. & Specht, M., 1997. *User Modeling and Adaptive Navigation Support in WWW-Based Tutoring Systems.* s.l., s.n.

Brusilovsky, P. & Millán, E., 2007. *The adaptive web: methods and strategies of web personalization. User models for adaptive hypermedia and adaptive educational systems.* Berlin: Springer-Verlag.

Belk, M. et al., 2015. The PAC Framework: Personalized Authentication and CAPTCHA based on Human Cognitive Factors.. *Technical Report, Department of Computer Science, University of Cyprus.*

Germanakos, P., Belk, M., Constantinides, A. & Samaras, G., 2015. *The PersonaWeb System: Personalizing E-Commerce Environments based on Human Factors.* s.l., s.n.

Constantinides, A., Belk, M., Germanakos, P. & Samaras, G., 2015. *The PersonaCheck System for Personalizing M-Commerce Checkout Processes.* Pittsburgh, IEEE.

Kleanthous, S. & Dimitrova, V., 2012. Adaptive notifications to support knowledge sharing in close-knit virtual communities. *User Modeling and User-Adapted Interaction,* 23(2-3), pp. 287-343.

Tsianos, N. et al., 2008. *User-Centric Profiling on the Basis of Cognitive and Emotional Characteristics: An Empirical Study.* Hannover, s.n.

Panayiotou, C. & Samaras, G., 2004. mPERSONA: Personalized portals for the wireless user: An agent approach. *Mobile Networks and Applications,* 9(6), pp. 663-667.

Papanikolaou, K. A., Grigoriadou, M., Kornilakis, H. & Magoulas, G. D., 2003. Personalizing the Interaction in a Web-based Educational Hypermedia System: the case of INSPIRE. *User Modeling and User-Adapted Interaction,* 13(3), p. 213–267.

Mitrovic, A. & Martin, B., 2002. *Evaluating the Effects of Open Student Models on Learning.* Malaga, s.n.

Anderson, C. R., Domingos, P. M. & Weld, D. S., 2001. *Personalizing web sites for mobile users.* New York, Association for Computing Machinery.

Maglio, P. P. & Barrett, R. C., 2000. Intermediaries personalize information streams. *Communications of the ACM,* 43(8), pp. 96-101.

Bogonikolos, N., Fragoudis, D. & Likothanassis, S., 1999. *ARCHIMIDES: An Intelligent Agent for Adaptive Personalized Navigation within a Web Server..* s.l., s.n.

Carro, R. M., Pulido, E. & Ródriguez, P., 1999. *TANGOW: Task-based Adaptive learNer Guidance On the WWW.* s.l., s.n.

Brusilovski, P., Kobsa, A. & Vassileva, J., 1998. *Adaptive Hypertext and Hypermedia.* 1 ed. s.l.:Springer Netherlands.

Bra, P. D. & Calvi, L., 1998. AHA! An open Adaptive Hypermedia Architecture. *New Review in Hypermedia and Multimedia ,* 4(1), pp. 115-140.

Neumann, G. & Zirvas, J., 1998. *SKILL - a Scalable Internet-Based Teaching and Learning System.* s.l., WebNet 98, World Conference on WWW.

Thomas, C. G. & Fischer, G., 1997. *Using agents to personalize the Web.* New York, Association for Computing Machinery.

datavisproject, n.d. *datavisproject.* [Online]
Available at: https://datavisproject.com
[Accessed 10 05 2020].

dataviscataloque, n.d. *dataviscataloque.* [Online]
Available at: https://dataviscataloque.com/search.html
[Accessed 10 05 2020].

Doumont, J.-L. & Vandenbroeck, P., 2002. Choosing the right graph. *IEEE Transactions on Professional Communication,* 45(1), pp. 1-6.

# Appendix 1 - Enterprise Platform Comparison

| Company | Software | Version | Price | On Premise | Cloud | Mobilty | API | Number of Data Sources |
|---|---|---|---|---|---|---|---|---|
| Microsoft | Power BI | April 2017 Update (2.45.4704.442) | 9.99 | Data can reside on-premises, but for sharing and collaboration the dashboards are stored in the Microsoft Azure cloud. | Yes | **Yes** | DAX/M,R | ~80 |
| Qlik | Qlik Sense QlikView | Qsense 3.1 (sep-2016) | 20 | Yes | Yes | Qlik Sense | Qlik Analytics Platform (QAP) | 71 |
| Tableau | Tableau | 10 | 42 | Yes | Yes | Yes | REST APIs and JavaScript | ~90 |
| SAS | SAS Visual Analytics (SAS BI) | 7.3 | Quote request | Yes | Yes | Yes | REST API | |
| SAP | SAP BusinessObjects Lumira and BusinessObjects Cloud | 4.2 | 185 | Yes | Yes | Yes - but needs improvement | REST API,Java | Limited |
| ThoughtSpot | ThoughtSpot | 6.1 | Quote request | Yes | Yes | Yes | REST API | |
| Oracle | Oracle Analytics Cloud | 5.5.4 | Quote request | Yes | Yes | Yes | REST API | Gartner 2017 - Highest Number of Combined Data Sources Compared to other Vendors |
| Sisense | Sisense | Cloud: 8.0.3 Windows: 8.2 | Quote request | Yes | Yes | Yes | REST API / Can also Extend Dashboard Functionality using JavaScript | High Number - ElastiCubes |
| Salesforce | Einstein Analytics | - | $150 / Month / User | No | Yes | Yes | REST API | - |
| YellowFin | YellowFin | 9.1 | Quote request | No | Yes | Yes | DashXML / SOAP | - |
| MicroStrategy | MicroStrategy | 2020x | Quote request | Yes | Yes | Yes | REST API | 70+ |
| TIBCO Software | TIBCO Spotfire | 10.7 | Quote request / Spotfire For Amazon Web Services starts at $0.99 per hour | Yes | Yes | Yes | REST API | 55 |
| IBM | Cognos Analytics | 11.1 | Starts at $15 / user / month | Yes | Yes | Yes | Promp API - JavaScript | High Number - Gartner 2017 |
| Looker - Google (Pending Acquisition) | Looker | 7.6 | Quote request | Yes | Yes | No | REST API | SQL Databases / No ETL process is included in this platform |
| Infor | Birst | 7.3 | Quote request | Yes | Yes | Yes | No | - |

| Company | Real Time Data and Dashboards | Complex Data Modeling | Custom Queries | Surrounding technologies | Expressions / Formulas | Update Schedule | Quick Insights | Community |
|---|---|---|---|---|---|---|---|---|
| Microsoft | Yes, Power BI is part of the Microsoft Data Platform and can stream data.<br><br>Azure Stream Analytics, IoT, predicted results from Machine Learning. | Yes | Yes, visual query editor +ribbon like in Excel allows to perform tasks such as:<br><br>- Connect to Data<br>- Shape and Combine Data<br>- Group Rows<br>- Pivot Columns<br>- Create Custom Columns<br>- Query Formulas | Tight integration with Microsoft ecosystem, supports Excel Based Add-ons (Power Query, Power Pivot, Power View and Power Map.<br><br>Good level of supporting features including alerts, print to pdf, etc. | DAX.<br><br>Consensus seems to be that DAX is the most powerful and versatile, with the added benefit that it is similar to Excel expressions. | Weekly (Online)<br><br>Monthly (Desktop) | Power BI lets you generate quick insights from any dataset and points out (in Natural Language form) insights such as correlations and outliers.<br><br>Qlik Narratives and Tableau Storytelling are NOT USPs and have nothing to do with Quick Insights. Also, Power BI has the same third party Narratives visual as Qlik | Power BI has a flourishing community, excellent documentation, and gives users the ability to suggest and vote for new features. |
| Qlik | Automatic refreshes - not "true" real time. | Yes | Yes, but uses SQL | Good integration with Office Suite to generate reports with NPrinting. Also offers scheduling of report distribution through email, and even publishing online. | Expression Editor | Every few months | No | Average forums, poor tech support according to Gartner 2017. |
| Tableau | No | Poor capabilities in combining data from different sources. Poor performance handling large and complex data has forced Tableau to plan to release a stand-alone data preparation tool (code-named Project Maestro) to address this issue. | Yes, but uses SQL | Many features are a work in progress, for example: event-based scheduling, conditional alerting, printing to PDF and PowerPoint, and collaboration and social platform integration are only available through partners, which adds to the TCO. | LOD Expressions (Level of Detail). | ~Semester + major update ever 1-2 years | No | Average forums. |
| SAS | No | Yes, one of the core strengths according to Gartner 2017. | Yes, but uses SQL and complicated windows and menus. | Requires additional product to integrate with Microsoft Office products. Poor or lacking functionality for print to PDF, scheduling reports. | Some degree of manipulation via expressions. | Last major update was in 2015, they're working on a new release for 2017. | No | Almost absent forums. |
| SAP | Yes | Yes | Both visual query builder and customer queries | Tight integration with SAP technologies and products such as Crystal Reports Enterprise,Crystal Reports 2016, Web inteligence | simple formulas, SQL Based expression, visual query builders | Varies, but usually long release cycles between major releases. A number of incremental SP are made available from 3-12 months accordinly | No | Average forums. |
| ThoughtSpot | Yes | Yes | TQL is the ThoughtSpot language for entering SQL commands. | SpotIQ - Augmented analytics capability. Can Live query a data warehouse (Amazon Redshift, Google BigQuery, and Azure Synapse) | ThoughtSpot's Formula Assistant | - | Yes using SpotIQ | Limited international presence, but one that is growing. |
| Oracle | Dashboards but Not Realtime | Yes - Above Average in Gartner 2017 | Yes, but uses SQL | Tight Integration with Oracle's ecosystem. Day by Day Mobile App. Oracle Analytics for Applications. Collaboration through Microsoft Teams and Slack. | Expression Editor | Every 1 to 2 Months | Oracle's Explain Feature NLG. Chatbot integration coupled with autogenerated insights. | Average forums. Thourough Documentation with Video Tutorials. |
| Sisense | Yes | Yes - ElastiCubes | Yes, but uses SQL | This product is mostly used as OEM / Embeded BI so it can be extended and also incorporated on any website or software product using it's endpoints. | Expression Editor - Similar to SQL Functions | Every 1 to 2 Months | Dashboard Widget Narratives NLG | Webinars and Workshops |
| Salesforce | Dashboards but Not Realtime | Strong Augmented Analytics Gartner 2020 | SAQL (Salesforce Analytics Query Language) | Einstein Prediction Builder, Sales Analytics, Service Analytics, Analytics Studio, Data Platform, Einstein Discovery and Einstein Data Insights. | Yes | Monthly Or Multiple Times a Month | Yes Shows Report Insight and Recomends Improvements | Very Big Community that is growing at 50% to 75% a year - Gartner 2020. Good Videos and Documentation / Forums. |
| YellowFin | Periodic Refresh - Not Realtime | Yes | FreeHand SQL | Can embed content from Tableau, Qlik (Qlik Sense) and Microsoft (Power BI). | FreeHand SQL | - | Yes | Small Community |
| MicroStrategy | Automatic refreshes - not "true" real time. | Give business users a gold-standard-like data exploration experience for very large and complex datasets and models. Gartner 2017. | Yes, but uses SQL | Can connect to Tableau, Power BI, and Qlik. | Yes | Platform release updates occure every three months / Platform releases every December | Yes - Also through HyperIntelligence | Conferences, Training, Online Tutorials / Documentation - Weekly Updates |
| TIBCO Software | Yes - Spotfire® Data Streams | Yes | Yes, language depending on the target data source | - | Yes | Every 2 to 3 Months | Intelligence engine identifies relationships in data and instantly recommends visualizations for lightning-fast insights. | Small Community |
| IBM | Yes - Cognos® Real-time Monitoring Dashboard | Yes - Star and Transactional Schema | SQL or MDX | - | Cognos® Transformer expression | Every 2 to 3 Months | Yes through NLG | Multiple Forums |
| Looker - Google (Pending Acquisition) | Yes | Data modeling requires coding | Yes - Writing SQL that then is directly executed on the underlying database | Gartner 2020 "Looker's key differentiator is native support for cloud-based analytic databases, particularly Amazon Redshift and Athena, Google BigQuery, Microsoft Azure and Snowflake" | LEXP | Monthly | No | Gartner 2020 - Positive reviews on availability and quality of partner resources and for its user community and training. |
| Infor | - | Yes - Drag and Drop and From multiple sources | - | - | - | - | Smart Insights | Not Much support can be found |

| Company | Custom Visuals & Download Gallery | Misc. | Visual Drill Down | Drag and Drop |
|---|---|---|---|---|
| Microsoft | Yes | Power BI is younger than competition and so has many lacking features and some are not implemented as well as they could be. For example, poor forecasting and no what-if scenarios, missing pivot tables, cannot show subtotals in the table visual, etc. | Average | Partial |
| Qlik | Yes, but no gallery | Qlik Sense, is the result of the competision pressure from companies such as MS and Tablaue in the cloud area. it provides a more user friendly self service interface for data exploration and visualization than QlikView. QlikView did not get dicommission, but sold in the enterprise users space | Excellent | Yes |
| Tableau | Yes | Good forecasting, what-if scenarios, good data interaction like highlighting data on a visual, removing certain elements temporarily, easy drilldown. | Good | Partial |
| SAS | Not custom visuals, only custom "graph objects". For example, a pie chart + bar chart in the same graph object. | Ugly interface and poor ease of use.<br><br>Disjointed product and workflow. SAS's BI capabilities are split among three products. | Very Good | Partial |
| SAP | No download gallery or market place, but capability of script based visuals | Gartner 2017 " Digital Boardroom is a differentiator: SAP's Digital Boardroom solution, which is built to be used with large touchscreen displays, has gained a lot of attention. It speaks well to the vision of a data-driven company and is particularly attractive to executives because it includes "what if" analysis and simulations. SAP can leverage its strategic position in a customer base of large enterprises and also protect its installed base against smaller vendors with less access to (and visibility with) senior executives." | Good | Extensive |
| ThoughtSpot | No | Primary interface for querying data is using NLP where a user can ask a question through speaking or typing. This tool does not cover the full requirements of an ABI and it requires third party applications for preparing and cleaning the data. Other than that it provides augmented analytics including the discovery of (anomalies, correlations) and also supports comparative analysis between data points without coding. | Good | No |
| Oracle | Only in DataVisualization Desktop | This tool is an end to end cloud solution, that has support for NLG that can be used with a number of languages. Moreover, it includes data management, infrastructure, analytics and analytic applications with focus on augmented analytics. Reports and dashboards with an integrated design experience are also included for interactive analysis. | Good | Yes |
| Sisense | Yes / There is a plugins library for extending Sisense functionality with visualizations, widgets etc | Gartner 2020 "Sisense provides an ABI platform that supports complex data projects by offering data preparation, analytics and visual exploration capabilities. Half of its ABI platform customers use the product in an OEM form". | Good | Yes |
| Salesforce | No | Gartner 2020 -"It remains strongest in terms of augmented analytics functionality. Einstein Analytics is much more likely to be embedded in business applications — commonly Salesforce's own apps — than other ABI platforms". | Good - Customizable | Yes |
| YellowFin | No | Gartner 2020 "Its capabilities span data preparation, Mode 1 reporting with scheduled distributions, Mode 2 visual exploration, and augmented analytics. All are accessed via a browser-based interface. Provides NLG natively and in a range of languages and supports data journalism. Yellowfin Signals, inform you about interesting changes in data". | Good | Yes |
| MicroStrategy | Yes / Can also extend with third party JS libraries | A comprehensive ABI platform with products for data visualization, advanced data connectivity and analytics. Moreover, it contains complementary mobile, cloud, embedded and identity analytics products. | Good | Yes |
| TIBCO Software | There is a download gallery with not much of visualizations / A framework JSVis is supported for creating custom visualizations with JS libraries | The platform supports dashboards with strong analytical capabilities, interactive visualizations, data preparation and augmented analytics. Moreover, the platform supports NLQ and NLG, it contains a mechanism that automatically suggests visualizations and also supports real-time screaming of data in many easily consumable forms. | Good | Yes |
| IBM | No | Supports the entire analytics life cycle from discovery to operationalization. Supports augmented analytics and supports statistically significant differences/insights, time series forecasting, key driver detection, NLP and NLG. Has reporting and visual exploration functionalities. Now Cognos also supports Planning Analytics. A variety of deployments is available covering all possible customer expectations. | Drill up or Drill down | Yes |
| Looker - Google (Pending Acquisition) | Yes - Users can extend Looker by downloading Applications, Models and Plugins. Plugins include visualizations that can be downloaded for extending the Looker's native visualization library | Looker is a modern ABI with reporting and dashboard functionality that is optimised for cloud databases as its performing operations in the actual database. Looker does not require in-memory storage optimizations as it performs operations on the actual data found in the database. LookML applies the business rules on the database. | Yes | Yes |
| Infor | No | End-to-end data warehouse, reporting and visualization platform built for the cloud and on Premises.Provides centralized and decentralized analytics, visual analytics and also includes smart insights. | Yes | Yes |

| Company | Bar Chart | Stack or Area chart | Line Chart | Combo Chart | Gantt Chart | Milestone trend analysis (MTA) | Radar Chart | Scatter Chart |
|---|---|---|---|---|---|---|---|---|
| Microsoft | Yes | Yes | Yes | Yes | Custom VIZ | No | Yes | Yes |
| Qlik | Yes | Yes | Yes | Yes | No (but workaround provided) | No | Yes | Yes |
| Tableau | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes |
| SAS | Yes | Yes | Yes | Limited | No (but workaround provided) | No | Yes | Yes |
| SAP | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| ThoughtSpot | Yes | yes | Yes | No | No | No | Yes | Yes |
| Oracle | Yes | Yes | Yes | Yes | No | No | Yes | Yes |
| Sisense | Yes | Yes | Yes | No | No | No | No | Yes |
| Salesforce | Yes | No | Yes | Yes | No | No | Yes | Yes |
| YellowFin | Yes | Yes | Yes | No | No | No | Yes | Yes |
| MicroStrategy | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes |
| TIBCO Software | Yes | Yes but with Workaround Provided | Yes | Yes | Yes but With Workaround | No | No | Yes |
| IBM | Yes | Yes | Yes | Yes | No | No | Yes | Yes |
| Looker - Google (Pending Acquisition) | Yes | Yes | Yes | Can Mix Types Together | No | No | Workaround in forums | Yes |
| Infor | Yes | Yes | Yes | No | No | No | No | No |

| Company | Grid Chart | Pie Chart | Polar Chart | Doughnut Chart | Block Chart or Heat map | Funnel Chart | Gauge Chart | Mekko Chart |
|---|---|---|---|---|---|---|---|---|
| Microsoft | Yes | Yes | No | Yes | Yes | Yes | Yes | Custom VIZ |
| Qlik | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Tableau | Text Table (Crosstab) Highlight Table | Yes | Trhough a workaround with Radar Chart | Yes | Heat Map | Yes | No | No (but workaround provided) |
| SAS | Yes | Yes | Trhiough Scripting | No | Yes | Limited | Yes | No (but workaround provided) |
| SAP | Basic | Yes | Yes | Yes | Yes | Yes | Speedometer | MarimekkoChar |
| ThoughtSpot | No | Yes | No | Their Bar Chart is shown as Doughnut Chart | No | Yes | No | No |
| Oracle | No | Yes | No | Yes | No | Yes | Yes | No |
| Sisense | No | Yes | Yes | No | No | Yes | Yes | No |
| Salesforce | No | Yes | No | Yes | Yes | Yes | Yes | No |
| YellowFin | Heat Grid | Yes | No | Ring Chart | Yes | Yes | No | No |
| MicroStrategy | Yes | Yes | Yes | Not included - Third Party Provider In Their Library | Yes | Yes | Yes - With third party DSK | Yes |
| TIBCO Software | No | Yes | No | No - Can with Visualization Framework JSVis | No | Heat Map Workaround | No | No |
| IBM | No | Yes | Yes | Yes - With Pie Workaround | Yes | No | Yes | Yes |
| Looker - Google (Pending Acquisition) | No | Yes | No | Yes | Offered in the library of custom visualizations | Yes | Yes | No |
| Infor | No | Yes | No | Yes | Yes | Yes | No | No |

| Company | Pivot Table | KPI Charts | Table/Matrix | Map | Bullet Graph | Histogram | KPI | TreeMap |
|---|---|---|---|---|---|---|---|---|
| Microsoft | Yes | No | Yes | Advance | No | No (but workaround provided) | Yes | Yes |
| Qlik | Yes | No | Yes | Yes | No | No | Yes | Yes |
| Tableau | No | Yes | Text Table (Crosstab) Highlight Table | Yes and also symbol map | Yes | Yes | Yes | Yes |
| SAS | Cross Tab | No (but workaround provided) | Yes | Yes | No (but workaround provided) | Yes | Yes | Yes |
| SAP | Yes, but limited interactivity | Yes | Spreadsheet | Yes, with a lot of options | Yes | Yes | Yes | Yes |
| ThoughtSpot | Yes | No | No | Yes | No | No | No | Yes |
| Oracle | Yes | No | Yes | Yes | No | No | No | Yes |
| Sisense | Yes | Yes | Yes | Yes | No | No | Yes | Yes |
| Salesforce | No | Yes | Yes | Yes | Yes | No | Yes | Yes |
| YellowFin | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes |
| MicroStrategy | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| TIBCO Software | No | Yes | Yes | Yes | Yes | Yes | Yes Using Bar Graph | Yes |
| IBM | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Looker - Google (Pending Acquisition) | Yes | Yes | Yes | Yes | No | Yes - With workaround | Yes | Yes |
| Infor | No | No | Yes | Yes | No | No | No | Yes |

| Company | Bubble chart | Packed Bubbles | Waterfall charts | Box-and-whisker Plot | Sankey Diagram | Network Diagram | Correlation Map | Decision Tree | Word/Text Map | Custom Visualization | Dashboard Concept |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Microsoft | Yes | Custom VIZ | Yes | Custom VIZ | Custom VIZ | Custom VIZ | No | Custom VIZ | Custom VIZ | Yes | Yes |
| Qlik | Yes | No | No | Yes | No | No | No | No | No | Limited | Yes |
| Tableau | Circle view | Yes | No | Yes | No | No | No | No | No | Limited | Yes |
| SAS | Yes | No (but workaround provided) | Yes | Limited | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| SAP | Yes | No | Delta Chart | Limited support through Candlestick chart | Extension | No | No | No | No | Yes | Yes |
| ThoughtSpot | Yes | No | Yes | No | Yes | No | No | No | No | Limited | Very Basic (Lacks Mapping Features) |
| Oracle | Yes | No | Yes | Yes | Yes | Yes | No | No | No | Yes | Yes |
| Sisense | No | No | No | Yes | No | No | No | No | No | Yes | Yes |
| Salesforce | Yes | No | Yes | No | Yes | No | No | No | No | Yes | Yes |
| YellowFin | Yes | No | Yes | Yes | No | No | No | No | No | Yes - Can Use third party JS SDK | Yes |
| MicroStrategy | Not included - Third Party Provider In Their Library | Not included - Third Party Provider In Their Library | Yes | Yes | Yes | Yes | No | No | Word Cloud | Yes - Can Use third party JS SDK | Yes |
| TIBCO Software | No - Can with Visualization Framework JSVis | No | Yes | Yes | No - Can with Visualization Framework JSVis | No - Can with Visualization Framework JSVis / Through their Library | No | Can Download through Library | No | Yes | Yes |
| IBM | Yes | Yes | Yes | No | No | Yes | No | No | Tag Cloud | Yes / Instructions Included | Yes |
| Looker - Google (Pending Acquisition) | Yes - Modification of a Scatter Plot | Offered in the library of custom visualizations | Yes | Yes | Yes | No | No | No | Yes | Yes can use D3 library | Yes |
| Infor | Yes | No | No | No | Yes | No | No | No | No | No - Can customize existing charts | Yes |

| Company | Predictive Analytics | Automated visualization suggestion | Natural Language Commands (Query) | Natural Language Audio Support | Visualization Description and Insights Feedback in natural language (natural language generation) | Other language based query or feedback Capabilities |
|---|---|---|---|---|---|---|
| Microsoft | Yes, but through Azure ML. Difficult to combine and link the data | Yes, based on the structure of the loaded from the data. But does not identify industry or business parterns | Yes, but limited dictionary of commands. No industry or context related features | Yes, through windows 10 clients with limited dictionary of commands. Cortana Analytics | No | Get answers based on questions including a fixed set of aggregate and calculation commands, feature names and visualization types |
| Qlik | No, but it can be done through extensive scripting with R | Limited and not industry or context related | No | No | Additional capabilities to describe the generated visualization - 'Narrative Science' a free extension | -- |
| Tableau | No, but it can be done through extensive scripting with R or integration with 3rd party platform such as SAS | Limited and not industry or context related | Yes, NLP is supported | No, but part of their roadmap | Wordsmith extension to describe the data and the visualizations | -- |
| SAS | As an addon directly to the platform (SAS Visual Statistics and SAS Analysis Server) | Limited and not industry or context related | No | No | No | -- |
| SAP | SAP BusinessObjects Predictive Analytics | Limited and not industry or context related | - | - | - | -- |
| ThoughtSpot | No, but it can be done through extensive scripting with R | Selects a chart based on the data of the search query not industry or context related | Yes - Main Feature | Yes - Main Feature | Yes - Main Feature | -- |
| Oracle | Yes | Yes, can be turned of. It selects a visualization based on the selected data elements. Not industry or context related | Yes | No | Yes - Language Narrative | - |
| Sisense | Yes - Not Used for very complex analysis | No - But incompatible visualizations to the data selected cannot be choosen | No - Has a CahtBot that can do basic actions from Skype, Slack and Messenger | No | Yes - Language Narrative | - |
| Salesforce | Yes - Einstein Predictive Builder | A number of suggested charts are shown depending on the selected data. Not industry or context related | Yes | - | Yes - Einstein Analytics Stories | - |
| YellowFin | Yes | Yes, by adding data a recommended visualization is displayed. Not industry or context related | - | - | Yes | - |
| MicroStrategy | Yes | Each visualization has data requirements that our selected data needs to adhere to. | Yes | - | Yes - Has Multiple Settings for each Narrative | - |
| TIBCO Software | Yes | Yes - Takes into account the relationships between data. | Yes | - | Yes | - |
| IBM | Yes - Predictive Models and Forecasts. | No | Yes | - | Yes | - |
| Looker - Google (Pending Acquisition) | Yes But Limited | No | No | No | No | - |
| Infor | Yes | Besed on selected data combinations | No | No | No - In plans | - |