Catalogue Visu: a Tool for Fast Visualization Prototyping

Jérémy Wambecke, Ivan Logre, Amandine Murgue

▶ **To cite this version:**

Jérémy Wambecke, Ivan Logre, Amandine Murgue. Catalogue Visu: a Tool for Fast Visualization Prototyping. 2023. hal-04014964

HAL Id: hal-04014964
https://hal.science/hal-04014964

Preprint submitted on 5 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Catalogue Visu : un Outil pour le Prototypage Rapide de Visualisations

Catalogue Visu: a Tool for Fast Visualization Prototyping

JÉRÉMY WAMBECKE, KSZ Lab, Kaizen Solutions, France
IVAN LOGRE, KSZ Lab, Kaizen Solutions, France
AMANDINE MURGUE, KSZ Lab, Kaizen Solutions, France

Fig. 1. Interface de *Catalogue Visu*. A gauche, la section pour choisir ses données. A droite, le choix de la visualisation. En haut, la configuration de la visualisation, visible au centre.

Visualization is an essential tool for exploring, understanding and presenting data. However, its use is limited by knowledge and tools. Novice users in visualization do not always know which visualization to choose for their data. The realization of a visualization with specific tools is often long and complex. In this article we present Catalogue Visu, an application allowing the free and fast realization of visualizations. Our main contribution is the design of the choice of the visualization type, which allows any user to choose an appropriate visualization for his needs. This method, the heart of Catalogue Visu, is iteratively validated by three user studies. Catalog Visu is accessible online and in permanent evolution, and is already used internally as a prototyping tool.

CCS Concepts: • **Human-centered computing** → **Visualization systems and tools**; **Empirical studies in interaction design**.

Additional Key Words and Phrases: Visualization design and evaluation methods, Interaction design

La visualisation est un outil indispensable pour explorer, comprendre et présenter les données. Toutefois son utilisation est limitée par la connaissance et les outils. Les utilisateurs novices en visualisation ne savent pas toujours quelle visualisation choisir pour leurs données. La réalisation d'une visualisation avec des outils spécifiques est souvent longue et complexe. Dans cet article nous présentons Catalogue Visu, une application permettant la réalisation libre et rapide de visualisations. Notre principale contribution est la conception du choix du type de visualisation, qui permet à tout type d'utilisateur de choisir une visualisation appropriée à ses besoins. Cette méthode, coeur de Catalogue Visu, est validée itérativement par trois études utilisateur. Catalogue Visu est accessible en ligne et en évolution permanente, et est déjà utilisé en interne comme outil de prototypage.

Mots-clés additionnels : Conception de visualisation, conception d'interaction

2023. Manuscript submitted to ACM

Reference:
This is the author's draft version of the work. It is posted here for your personal use. Not for distribution. The definitive version of record was published in IHM '23: Proceedings of the 34th Conference on l'Interaction Humain-Machine.
ISBN 978-1-4503-9825-1
https://doi.org/10.1145/3583961.3583969

1 INTRODUCTION

La visualisation est beaucoup utilisée pour comprendre et raisonner sur les données et leurs caractéristiques, en traduisant des valeurs brutes en une représentation visuelle [5]. Elle permet aux données d'être interprétées en tant qu'information en se basant sur la puissance du système visuel humain. Avec l'augmentation des flux de données du web et la répartition croissante des capteurs personnels, par exemple les montres connectées, les utilisateurs ont de plus en plus besoin de visualisation. Cette augmentation de l'utilisation entraîne une augmentation de l'hétérogénéité des données, qui elle même entraîne une augmentation des types de visualisation. Ainsi, au delà des représentations "classiques" telles que diagramme en bâtons, en courbe ou camembert, des visualisations spécialisées sont créées et partagées en ligne, par exemple pour les données sportives [15] ou encore les résultats de vote [4]. Cette diversification permet de nouvelles interprétations des caractéristiques des données étudiées.

Avec cette démocratisation des données, nous constatons également une diversification des profils d'utilisateurs. Tout d'abord réservée aux experts métier, la visualisation est maintenant utilisée par des utilisateurs de profils variés, pour des besoins personnels ou professionnels [11, 12]. Toutefois les utilisateurs ont des besoins et des capacités de compréhension de la visualisation différents [16]. Ainsi il n'est pas facile pour eux de choisir quelle visualisation sera pertinente en fonction des données et de l'objectif de l'analyse. Dans l'état de l'art, les profils différents des utilisateurs selon la compréhension de visualisations sont rarement pris en compte. De plus les plate-formes de création de visualisations actuelles sont soit limitées en termes de création, soient complexes à apprendre.

Dans cet article, nous présentons *Catalogue Visu*, une plate-forme en ligne permettant aux utilisateurs de créer des visualisations avec une approche exploratoire. *Catalogue Visu* couvre deux principaux cas d'utilisation : (1) la formation à la visualisation ainsi que la découverte de nouveaux types de visualisation, et (2) le prototypage de visualisations spécifiques destinées à des applications métier. Les visualisations sont réalisées par trois étapes : choix des données, choix du type de visualisation et association des données aux variables visuelles. *Catalogue Visu* est conçu pour permettre de tester des types de visualisation rapidement, soit en partant de données d'exemple illustrant les visualisations, soit en important ses propres données.

Le choix du type de visualisation est crucial car il détermine la façon dont le système visuel de l'audience va être utilisé, et donc l'efficacité de la visualisation. Cette sélection est basée sur une taxonomie des types de visualisation selon des *Insights* [7, 9, 21, 22, 24], qui peuvent être définis comme les caractéristiques des données mises en avant par l'analyse. Nous avons conçu itérativement cette étape en réalisant trois études utilisateur, permettant respectivement de représenter les objectifs des utilisateurs en un ensemble d'*Insights*, de représenter graphiquement ceux-ci puis enfin de valider l'étape de sélection de la visualisation.

Nos contributions sont (1) la conception d'un choix du type de visualisation basé sur une taxonomie liant les types de visualisation à des *Insights* et (2) la conception de *Catalogue Visu*, une application web appliquant ce choix à la réalisation de visualisations.

Nous commençons par présenter l'état de l'art dans la Section 2. Nous présentons ensuite les motivations nous ayant conduit à la conception de *Catalogue Visu* dans la Section 3. Le choix de la visualisation ainsi que la taxonomie

permettant cette étape sont ensuite présentés en Section 4. La conception et l'implémentation de *Catalogue Visu* suivent en Section 5. Enfin nous présentons les travaux futurs en Section 6 avant de conclure en Section 7.

2 ÉTAT DE L'ART

2.1 *Insights*

Card *et al.* [6] affirment que le but principal de la visualisation est de fournir des *Insights*. Dans la littérature, plusieurs définitions d'*Insight* ont été proposées [7, 9, 21, 22, 24]. Ce concept est toutefois vague, abstrait et aucune définition ne fait consensus. D'après ces différentes définitions, nous pouvons définir les *Insights* comme des enseignements tirés des données. Mais les *Insights* peuvent également être vus comme des objectifs de la visualisation, en définissant ce que l'utilisateur veut en tirer. Ainsi *Catalogue Visu* est basé sur ce concept d'*Insights*, vus comme des objectifs, afin de catégoriser les visualisations selon ce qu'elles peuvent apporter aux utilisateurs.

2.2 Tâches de visualisation et taxonomies

Plusieurs taxonomies de visualisations ont été présentées dans la littérature. Schulz et al. [25] proposent de définir la tâche de visualisation avec six dimensions. Deux sont particulièrement liées à notre contribution : *pourquoi* qui définit les objectifs de la tâche et *quoi* qui définit les caractéristiques des données étudiées. Le *DataViz Catalogue* [23] définit une terminologie de fonctions visuelles basée sur les connaissances extraites à partir de la communauté journalistique, afin d'établir des liens entre les visualisations et les fonctions visuelles. Cependant cette taxonomie reste théorique, et aucun outil ne permet de générer des visualisations à partir de la taxonomie. Tory et Möller [27] ont présenté une taxonomie pour les experts en visualisation, basée sur le modèle de données et le modèle d'utilisateur, qui se base sur deux critères : le type des données (continu ou discret) et les canaux visuels utilisés. Plusieurs contributions de Schneiderman [26] et de *the Data Visualization Service of the Duke University Library* [31] ont présenté une catégorisation des visualisations en étendant la taxonomie de Tory et Möller [27]. Dans ces contributions, les visualisations sont séparées en catégories disjointes se basant sur leur définition de *dimension*. Cette catégorisation peut toutefois être peu parlante pour l'utilisateur. Chen et al. [8] ont présenté une taxonomie d'*Insights* dépendant du domaine. Cependant les termes de la taxonomie ne sont pas explicitement liés aux visualisations. Vázquez-Ingelmo et al. [28] ont présenté un méta-modèle de *dashboard* composé de trois principaux éléments : modélisation des utilisateurs, conception de l'agencement et abstraction des visualisations. Ce modèle est théorique, ainsi la génération de visualisations et la validation du modèle ne sont pas abordés. Choe et al. [9] ont étudié la réalisation et l'utilisation de visualisations créées par des utilisateurs non experts, pour analyser leurs propres données personnelles. Ils en ont déduit une taxonomie basée sur 8 types d'*Insights* qui peuvent être tirés de ces visualisations. Toutefois l'association entre visualisations et *Insights* n'est pas abordée par cet article.

2.3 Suggestions de visualisations

Dans les travaux précédents, plusieurs solutions ont été présentées afin de guider l'utilisateur dans la réalisation de la visualisation. Une approche permettant aux utilisateurs de caractériser leurs besoins en visualisation a été présentée par Logre et Déry [17]. L'évaluation de cette approche montre que les participants ont été capables d'utiliser des concepts haut niveau pour exprimer leurs besoins de visualisation et évaluer les visualisations qui satisfont ces *Insights*. L'expérimentation a été effectuée uniquement sur la base d'entretiens, mais aucun prototype n'a été utilisé. Nous nous

inspirons de ce protocole dans notre expérimentation, mais en utilisant un prototype fonctionnel. *VizAssist* [3][1] permet de guider l'utilisateur sur plusieurs étapes de conception de la visualisation. Après avoir importé leurs données, les utilisateurs peuvent sélectionner des tâches de fouille de données à appliquer. Des filtres sur les données peuvent être ajoutés, et plusieurs visualisations possibles sont suggérées. L'utilité de *VizAssist* est évaluée en comparaison de *VRMiner* [1], un système similaire des mêmes auteurs. *Catalogue Visu* est proche de *VizAssist* mais est moins restrictif car il permet de guider l'utilisateur dans le choix de visualisation. VizDeck [14] est une application web pour l'analyse visuelle exploratoire de données relationnelles non organisées. Dans cet outil, la responsabilité du choix d'une visualisation pertinente est laissée à l'utilisateur avec peu de guidage. L'objectif est d'avantage porté sur la détection automatique de caractéristiques statistiques et sur la visualisation de ces caractéristiques. VizRec [20] est un système de recommandation de visualisations en deux étapes : *(i)* un système basé sur des règles génère des visualisations appropriées aux données spécifiques, puis *(ii)* le résultat est trié et filtré selon des préférences utilisateur. Toutefois la tâche de l'utilisateur n'est pas prise en compte, contrairement à notre approche qui place plus d'importance dans la tâche que dans les données. Kaur and Owonibi [13] rassemblent les stratégies de recommandation de visualisations afin de rassembler les ressemblances et pointer les lacunes. Leur contribution est présentée comme une grille, définissant quatre familles de recommandations : caractéristiques des données, de la tâche, orienté par les connaissances métiers ou par les préférences de l'utilisateur. Many Eyes [29] est un système de guidage pour la visualisation conçu pour les utilisateurs novices. Il permet de sélectionner des colonnes de données pertinentes parmi des jeux de données. Le système propose alors des visualisations correspondant à ce nouveau jeu de données ainsi construit. Enfin Voyager [30] est une application web utilisation le système de recommandation Compass[2] afin de suggérer des types de visualisations à partir des caractéristiques statistiques des données.

3 MOTIVATIONS ET CAS D'UTILISATION

Dans cette section, nous présentons les motivations nous ayant conduit à concevoir *Catalogue Visu*, et en particulier la méthode utilisée pour choisir le type de visualisation. *Catalogue Visu* couvre plusieurs cas d'utilisations : la formation à la visualisation ; la découverte et la création de visualisations par des utilisateurs novices ; la découverte de nouvelles visualisations par des utilisateurs expérimentés ; le prototypage de visualisations spécifiques destinées à des applications métier.

Formation à la visualisation. Les utilisateurs novices en visualisation ont besoin d'une aide pour choisir le type de visualisation. En utilisant des outils "classiques" et faciles d'accès, il y a un risque que les visualisations soient simples et non optimales, voire peu efficaces [9]. D'un autre côté, l'utilisation d'outils tels que Tableau[3] demande une courbe d'apprentissage qui peut rebuter les utilisateurs. Même pour les utilisateurs confirmés en visualisation, *Catalogue Visu* permettrait non seulement de découvrir de nouveaux types de visualisation, mais également de tester rapidement ces types de visualisation sur leurs données.

L'exploration étant importante pour l'apprentissage, *Catalogue Visu* doit pouvoir permettre facilement l'ajustement des données utilisées : en démarrant avec des données d'exemple adaptées à la visualisation, puis en important ses propres données. L'association entre les données et les variables visuelles doit également être facilement modifiable. L'essai d'associations différentes avec les mêmes données et avec un rendu de la visualisation immédiat permet aux utilisateurs de bien comprendre le fonctionnement de la visualisation et ses variables visuelles. Dans ce contexte

[1] http://www.vizassist.fr/
[2] https://github.com/vega/compass
[3] https ://www.tableau.com/

d'exploration, les utilisateurs devraient pouvoir commencer par n'importe quelle étape. Soit importer des données ou choisir des données d'exemple, puis déterminer ensuite quelle visualisation conviendrait ; soit choisir un type de visualisation inconnu ou attirant l'utilisateur, puis utiliser des données d'exemple spécifiques pour analyser son utilisation ; soit importer une visualisation déjà réalisée pour la modifier.

Prototypage de visualisations spécifiques. Pour les applications métier, le prototypage des visualisations est indispensable dans la réalisation d'applications de visualisation. Les utilisateurs qui utilisent la visualisation dans le cadre d'un domaine métier (par exemple l'énergie [10]) ont besoin de visualisations spécifiques, mais qui se basent sur des types de visualisation adaptés pour ce besoin. Lors de la réalisation d'une nouvelle application de visualisation, une plate-forme de prototype de visualisation est un outil très utile pour les concepteurs. Les designers UX récoltent les besoins métier pour l'application, puis présentent à l'aide de l'outil des visualisations de différents types qui pourraient être adaptées aux besoins. L'outil sert alors de plate-forme de présentation des visualisations, et permet aux designers d'expliquer le fonctionnement des visualisations aux acteurs métier. Cette présentation est réalisée soit avec des données d'exemple, soit directement avec les données métier dans la plate-forme.

4 CHOIX DU TYPE DE VISUALISATION

Dans cette section, nous présentons la conception itérative du choix du type de visualisation. Cette étape est la plus importante de *Catalogue Visu* : elle va déterminer le type de visualisation choisi par l'utilisateur en fonction de ses besoins. Pour cela nous définissons une taxonomie associant les *Insights* (voir Section 2) aux types de visualisations. Ces *Insights* permettent à l'utilisateur de définir ses besoins, en filtrant les visualisations disponibles selon un ensemble d'*Insights* choisi.

4.1 Taxonomie d'*Insights*

Utiliser les variables visuelles [2, 18, 19] tel que le font les concepteurs de visualisations nécessite une expertise en visualisation et perception humaine qui n'est pas atteignable par des utilisateurs novices. Ainsi dans *Catalogue Visu* le choix de visualisation est basé sur une représentation abstraite des besoins, compréhensible sans pré-requis de connaissances. Lors de cette étape, la réflexion doit se porter au niveau du besoin plutôt que des données. Les utilisateurs doivent choisir le type de visualisation à utiliser sans réfléchir au résultat final (par exemple quelles variables graphiques utiliser) mais plutôt en ciblant le besoin que doit remplir la visualisation. Pour cela nous devons proposer le niveau d'abstraction adéquat aux utilisateurs. Dans notre cas, nous avons choisi de nous baser sur les *Insights* (voir Section 2.1). Un *Insight* peut être vu comme une idée ou une caractéristique qui est issue de l'analyse de la visualisation.

Nous sommes partis de l'ensemble des *Insights* du *DataViz Catalogue* [23]. Nous avons conduit des entretiens avec des utilisateurs potentiels ayant des compétences en visualisation variées, mais qui ont besoin d'utiliser la visualisation au niveau professionnel ou personnel. Nous avons recruté 8 participante(s) pour ces entretiens, de domaines gestion / administration (2), analyse de données (2), développement informatique (2), UX (2). Les participant(e)s avaient entre 1 et 8 ans d'expérience professionnelle. Les entretiens se sont déroulés sous forme de deux questions ouvertes sur (1) l'utilisation actuelle de la visualisation et sur (2) une utilisation potentielle de *Catalogue Visu*. Pour cette deuxième question nous présentons les *Insights* du *DataViz Catalogue*. A la suite de ces entretiens, nous avons décidé de retirer les *Insights How things work*, *Process & Methods*, *Concepts*, *Analyzing text* and *Reference tool* qui s'apparentent plus aux schémas qu'à la visualisation de données. Les réponses des participants à la seconde question ont montré que certaines tâches ne sont pas couvertes par les *Insights* existants : analyser les variations des données, agglomérer ou discrétiser

Insight	Description	Exemples de visualisations associées	
Comparison	Comparaison de données	100% Stacked Area Chart	100% Stacked Bar Chart
Discrete	Application d'une discrétisation des données	Bar Chart	Histogram
Distribution	Répartition des données selon une ou plusieurs variables	Density Plot	Pie Chart
Extremum	Minimum et maximum des données	Candlestick Chart	Gauge
Flow	Flux de données	Flow Map	Parallel Sets
Hierarchy	Hiérarchisation des données sur plusieurs niveaux	Sunburst Diagram	Treemap
Location	Positions géographiques ou géométriques	Bubble Map	Dot Map
Over time	Evolution des données au cours du temps	Line Chart	Spiral Plot
Part of a whole	Inclusion de données dans un ensemble	Circle Packing	Marimekko Chart
Pattern	Répétition de structures dans les données	Heatmap	Stream Graph
Proportion	Comparaison des données sous forme d'aires	Stacked Area Chart	Stacked Bar Chart
Range	Intervalles des données	Span Chart	Bars Whisker Plot
Relationship	Relations entre données	Arc Diagram	Network Diagram
Variation	Variation selon une variable, pas forcément temporelle	Multiset Line Chart	Stacked Area Chart

FIG. 2. Ensemble des *Insights* utilisés pour spécifier le choix du type de visualisation avec leur description. Seules deux visualisations sont données en exemple pour chaque *Insight* par souci de place, mais un *Insight* peut être associé à plus de visualisations.

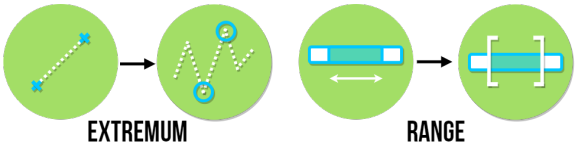

FIG. 3. Modification des représentations des *Insights Extremum* et *Range* après l'étude utilisateur initiale.

les données et trouver les minimum et maximum dans les données. Nous avons donc décidé d'ajouter trois nouveaux *Insights* qui expriment des besoins non présents dans la taxonomie initiale : *Variations*, *Discretization* et *Extremum*. La Figure 2 représente l'ensemble des 14 *Insights* ainsi obtenu.

Dans *Catalogue Visu*, le choix des types de visualisation est réalisé via une méthode de filtrage par *Insights*. Tous les *Insights* et les types de visualisation sont présentés. Lorsque l'utilisateur sélectionne un *Insight*, seuls les types de visualisation correspondant sont affichés. L'utilisateur peut alors sélectionner un ensemble d'*Insights*, correspondant à ses besoins, afin de sélectionner une visualisation parmi celles qui correspondent. Certaines associations d'*Insights* ne correspondent à aucune visualisation : l'utilisateur peut alors déselectionner des *Insights*. Il est aussi possible de sélectionner directement un type de visualisation sans spécifier d'*Insights*, par exemple dans le cas où l'utilisateur sait à priori quel type de visualisation il veut utiliser.

4.2 Représentation des *Insights*

Afin de simplifier visuellement l'étape de choix de la visualisation, les *Insights* sont représentés par des icônes car le nombre d'*Insights* est trop important pour proposer une description textuelle en permanence. Nous avons mené une étude utilisateur pour évaluer l'association entre les *Insights* et leurs icônes. L'objectif est de valider que les utilisateurs peuvent reconnaître les *Insights* à partir de leurs icônes, et ainsi permettre une caractérisation rapide du besoin par les utilisateurs.

Dans cette étude les participants doivent lier les icônes des *Insights* à leur description. Les icônes des 14 *Insights* sont présentés aux participants dans un ordre aléatoire. Pour chaque icône, les participants doivent choisir à l'aide de boutons radio à quel nom d'*Insight* elle correspond. Les participants peuvent également sélectionner "autre" s'ils pensent qu'aucune description ne correspond à l'icône. Les noms sont toujours tous affichés, de sorte qu'un même nom peut être choisi plusieurs fois pour une même icône. A titre d'étude exploratoire additionnelle, pour chaque icône présentée les utilisateurs peuvent saisir dans un champ de texte des exemples de type de visualisations correspondant à cet *Insight*.

Dans un premier temps nous avons conduit une étude pilote avec 4 participants. Les icônes des *Insights Extremum* et *Range* ont été jugées problématiques par les 4 participants. Nous avons alors conçues de nouvelles versions de ces icônes, que nous avons présentées à ces participants (Figure 3). Ces nouvelles versions ont été validées par les 4 participants et sont alors devenues les nouvelles versions des icônes pour l'étude.

Nous avons ensuite mené l'étude en ligne avec 113 participants. Les utilisateurs devaient auto-évaluer leur niveau de visualisation parmi *Beginner, Novice, Supporter, Expert or Ambassador*. Ils devaient également spécifier leur activité professionnelle parmi industriel (ingénieur, manager...), enseignant / chercheur, étudiant ou autre (avec un texte libre). Enfin ils pouvaient remplir de manière optionnelle leur genre et âge. Les activités professionnelles des participants sont réparties dans toutes les catégories, y compris *autre* avec par exemple artisan et psychologue. La distribution du niveau de visualisation des participants est centrée sur la valeur centrale (*supporter*). L'âge médian est de 26 ans, l'âge moyen

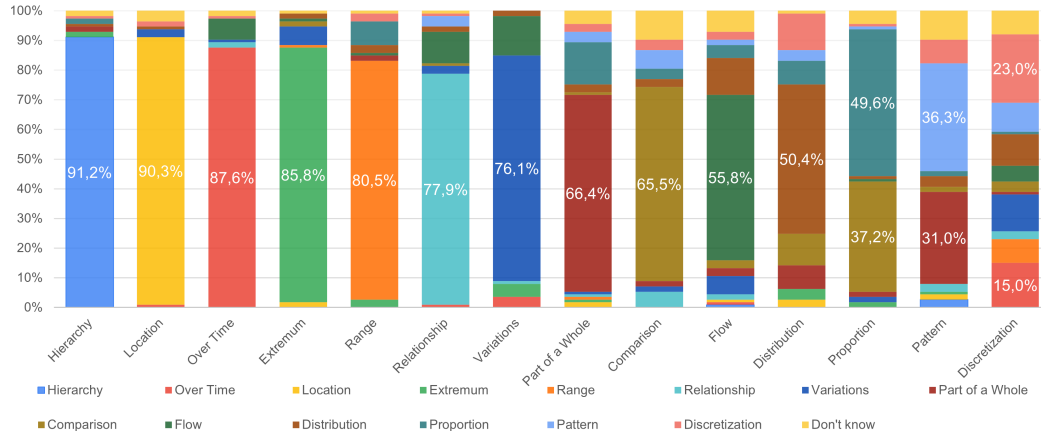

Fig. 4. Résultats de l'étude utilisateur sur la représentation des *Insights*. Chaque icône d'*Insight* est représentée par un bâton, représentant la proportion de chaque réponse possible : un nom d'*Insight* ou "autre".

de 30.3 ans avec une déviation standard de 10 ans. 23% des participants ayant indiqué leur genre sont des femmes, contre 77% d'hommes. Les résultats sont présentés sur la Figure Figure 4.

Pour chaque icône, le nom d'*Insight* ayant le plus de réponses est le nom attendu. Nous pouvons toutefois constater une grande disparité pour certaines icônes. En raison de la diversité des profils des participants, qui reflète la diversité des utilisateurs cibles, il serait inutile de fixer un seuil d'acceptation des résultats. Nous pouvons cependant noter que les trois derniers *Insights* ont présenté des problèmes aux participants.

L'icône de *Proportion* a été confondue avec l'*Insight Comparaison*. Pour l'icône de *Proportion*, environ la moitié des participants (49.6%) a donné la réponse attendue, mais un nombre important d'entre eux (37.2%) ont répondu *Comparaison*. Nous pensons que c'est en raison de la définition de *Proportion* qui est *une comparaison des données représentées avec des aires proportionnelles à leur valeur*. Il est intéressant de constater que cette confusion n'est pas bidirectionnelle, car seulement 3.5% des participants ont répondu *Proportion* lorsque l'icône *Comparison* est présentée. L'icône *Pattern* est reconnue correctement par seulement 36.3% des participants, et une grande partie des participants la confond avec *Part of a whole*. L'icône *Discretization* présente les plus mauvais résultats. Elle présente une grande disparité de réponses différentes, comme nous pouvons le voir sur sa barre dans la Figure 4. Au delà de la représentation de l'icône, le concept de discrétisation vient du domaine des données numériques et peut ne pas être compris par l'ensemble des participants en raison de leur variété de profils. Cela indiquerait un besoin de repenser l'*Insight* lui-même pour le rendre plus compréhensible, de même que son opposé *Variations* qui traduit une visualisation de données continues. L'*Insight Discretization* est toutefois conservé dans la taxonomie car il permet d'exprimer un besoin de discrétiser des données pour les visualiser, par exemple en les agrégeant.

4.3 Évaluation de la taxonomie

Après avoir défini l'ensemble des *Insights* et leur représentation par des icônes, nous avons évalué la pertinence de la taxonomie, c'est à dire l'association entre les *Insights* et les types de visualisations. Le but de l'évaluation est d'identifier les relations pouvant poser des problèmes aux utilisateurs et essayer d'y apporter des solutions. Atteindre une satisfaction du choix de visualisation de 100% serait illusoire, car les relations entre *Insights* et types de visualisations dépendent du

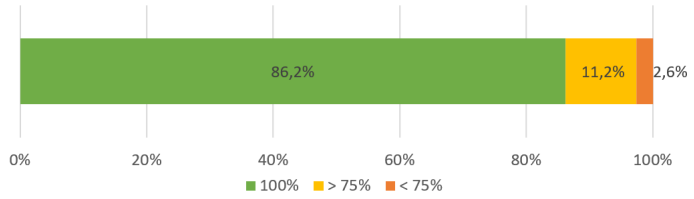

FIG. 5. Résultats de l'étude sur la caractérisation des types de visualisation par les *Insights*

domaine d'application. Les *Insights* pourraient même être adaptés au domaine. Ainsi nous plaçons arbitrairement un seuil d'acceptation de 75% pour les relations entre *Insights* et types de visualisation.

Pour cette évaluation nous avons conduit des entretiens avec des participants ayant des compétences en visualisation et qui utilisent la visualisation régulièrement. 8 participants (4 hommes et 4 femmes) ont ainsi été recrutés, de différents domaines : IHM, architecture logicielle, biologie, mathématiques appliquées et géologie. Ces participants étaient des ingénieurs expérimentés ou des professeurs d'université.

L'étude se déroule sur l'interface de choix de visualisation uniquement (Figure 8). La première étape est la présentation des *Insights*. Après avoir présenté cette interface, nous présentons aux participants chaque *Insight* ainsi que leurs définitions. Les participants peuvent alors discuter de la définition et de la signification des *Insights*. La seconde étape est la présentation de l'association des *Insights* avec les types de visualisations. Chaque *Insight* est sélectionné par l'expérimentateur, et *Catalogue Visu* affiche les types de visualisations associés. Le participant commente alors les types de visualisation associés à l'*Insight* sélectionné. Finalement le participant indique si la relation est appropriée (oui / non / je ne sais pas). Afin de limiter le temps de l'étude, chaque participant ne commente que 7 *Insights* sur 14. Toutefois les ensembles d'*Insights* sélectionnés sont répartis de sorte que chaque relation *Insight* / type de visualisation est évaluée le même nombre de fois par les experts.

Les données résultantes de ces entretiens correspondent à un pourcentage de validation pour chaque relation *Insight* / type de visualisation, présenté sur la Figure 5. Nous interprétons les résultats en trois groupes :

- 86.2% des associations ont un score de 100% de validation (tous les participants ont indiqué que la visualisation satisfait le besoin).
- 11.2% des relations ont été validées par 3 des 4 experts.
- 2.6% des relations ont un résultat plus bas et sont jugées problématiques.

Les 5 relations jugées problématiques sont *(visualisation ↔ insight)* : Parallel sets ↔ Flows, Flow Map ↔ Distribution, Flow Map ↔ Discretization, Gauge ↔ Discretization et Radar ↔ Patterns. Pour améliorer ces cas, nous prévoyons comme travaux futurs des groupes de travail avec des concepteurs de visualisations. Nous prévoyons également d'augmenter l'expressivité de la relation entre *Insights* et types de visualisations. Cette relation pourrait par exemple être pondérée, ce qui permettrait d'exprimer des nuances dans les associations.

4.4 Évaluation du choix de visualisation

Afin de valider les capacités du *Catalogue Visu* à permettre à des utilisateurs de générer des visualisations correspondant à leurs besoins d'analyse, nous avons mené une expérimentation avec des participants. Cette section présente le protocole, les résultats et les conclusions de cette étude.

Scénario	Manuel (s)	Taxonomie (s)	Gain (%)
Scénario 1	192	43	77.6
Scénario 2	273	69	74.9
Scénario 3	280	67	76.0
Scénario 4	290	82	71.7
Moyenne	259	65	74.8

Table 1. Résultats de l'étude sur le choix du type de visualisation : temps moyen (s) pour chaque condition et gain (%), par scénario

Le protocole de l'expérimentation est inspiré de Logre and Déry [17], car ce dernier évaluait un choix de visualisation en fonction de besoins utilisateur et est donc proche de notre expérimentation. Le but de l'expérimentation est de comparer le gain obtenu pour sélectionner des types de visualisation en utilisant le système de filtrage par *Insights*, par rapport à une sélection classique parmi tous les types de visualisation.

Ce choix est effectué sur quatre scénarios différents, basés sur des données de course en montagne :

(1) Visualiser la durée d'entraînement cumulée par semaine sur une année, pour identifier des périodes de relâchement et déduire des motifs dans les données, par exemple des périodes d'activité réduite après des périodes d'activité intensive.
(2) Mettre en relation la période de l'année avec la durée d'entraînement et la vitesse, pour analyser un éventuel impact saisonnier (froid ou chaleur) sur la performance.
(3) Comparer la vitesse et l'altitude au cours du temps, pour identifier si les diminutions de vitesse sont dues à l'inclinaison du sol ou à la fatigue.
(4) Analyser l'impact des variations d'altitude (et non juste de l'altitude, contrairement au scénario précédent) sur la vitesse, et ainsi l'effort nécessaire pour différentes parties de la course.

Deux tâches sont effectuées par les participants :

- **Manuel :** tous les types de visualisations sont présentés aux participants, qui doivent en choisir un sous-ensemble qui conviendrait pour générer une visualisation correspondant au scénario.
- **Taxonomie :** Les participants utilisent le système de filtrage par *Insights* pour sélectionner un sous-ensemble de types de visualisations qui correspondraient au scénario. Cette condition correspond à une utilisation standard de *Catalogue Visu*, avec sa fonctionnalité de filtrage des types de visualisation.

Dans les deux tâches, aucun temps n'est pré-défini et la tâche est considérée comme terminée lorsque le participant est satisfait des types de visualisations choisies.

Nous avons recruté 6 participants, qui ont déclaré ne pas avoir de connaissances en conception de visualisations. Au début de l'expérimentation, les participants sont familiarisés pendant 10 minutes aux *Insights* et aux types de visualisation disponibles. La moitié des participants commencent par la condition **Manuel**, l'autre moitié commencent par l'autre condition **Taxonomie**.

Les résultats sont présentés dans la Table 1. Pour chaque scénario, nous présentons le temps moyen passé pour choisir les types de visualisation pour chaque condition. Nous présentons également la différence de temps (gain) entre la condition **Manuel** et la condition **Taxonomie**. Nous pouvons constater que pour chaque scénario les participants passent moins de temps avec la condition **Taxonomie**, avec un gain moyen de 74.8%. Ces résultats confirment que le filtrage par *Insights* permet aux utilisateurs n'ayant pas de connaissances en visualisation de choisir plus rapidement

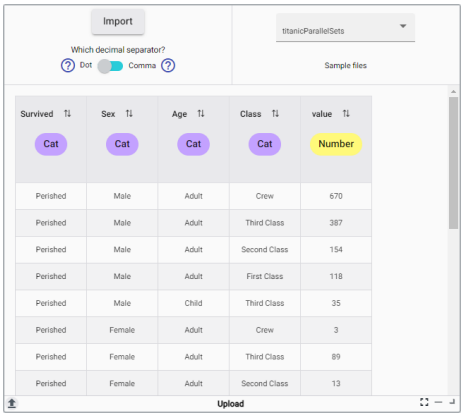

FIG. 6. Fenêtre d'importation des données.

des types de visualisation qui conviendraient pour un scénario donné, par rapport à une sélection parmi tous les types de visualisation.

5 *CATALOGUE VISU* : UNE PLATE-FORME DE PROTOTYPAGE DE VISUALISATIONS

Afin de mettre en pratique la sélection de visualisation ainsi conçue, nous avons développé *Catalogue Visu*, un outil permettant la réalisation simple et rapide de visualisations sous forme de prototypes. Nous présentons ici sa conception et son implémentation.

5.1 Organisation de l'interface

L'interface de *Catalogue Visu* est basée sur quatre zones principales (Figure 1) : le choix des données (*upload*), la sélection de visualisation (*selection*, voir Section 2.1), le choix des variables visuelles utilisées (*mapping*) et la visualisation. Ces zones sont représentées sous forme de fenêtres agençables librement par les utilisateurs : leurs tailles et position sont ajustables ; elles peuvent être mises en plein écran pour se concentrer sur une étape ; elles peuvent être cachées pour n'en garder que certaines. Cette liberté d'organisation des étapes permet de couvrir différents cas d'utilisation. Par exemple en début de conception, l'utilisateur peut se concentrer sur le choix de visualisation qui devra alors être central et occuper de la place. A la fin de la conception, la partie du rendu de la visualisation peut être maximisée pour l'analyse du résultat. Il est possible de ne conserver que le rendu de la visualisation et l'étape d'association, pour explorer différentes associations avec un même jeu de données et un même type de visualisation. Toutefois cette liberté apporte également un risque que l'utilisateur soit perdu dans l'utilisation des zones, surtout lors de la première découverte. Des agencements par défaut sont donc proposés pour organiser les étapes automatiquement selon des cas d'utilisation : équilibre entre les étapes, concentration sur le choix de la visualisation, concentration sur les données importées et concentration sur le rendu de la visualisation.

5.2 Choix des données

La Figure 6 présente la fenêtre de choix des données à visualiser. Deux possibilités sont offertes aux utilisateurs pour choisir des données : soit importer leurs propres données depuis un fichier (zone en haut à gauche), soit utiliser un des jeux de données d'exemple fournis (zone en haut à droite). Les jeux de données d'exemple sont spécifiquement

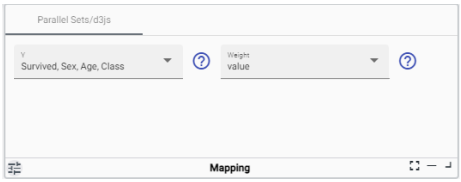

FIG. 7. Fenêtre de choix des associations entre données et variables visuelles.

FIG. 8. Fenêtre de sélection de la visualisation.

choisis pour démontrer l'utilisation des différents types de visualisation. Actuellement il est possible d'utiliser des jeux de données au format csv ou json.

5.3 Choix de la visualisation

La Figure 8 présente la fenêtre de choix de la visualisation. Cette fenêtre permet de choisir une visualisation en utilisant un système de filtrage par *Insights*. Au début, toutes les visualisations sont affichées. Les *Insights* sont tous présentés en haut de la fenêtre. Lorsqu'un utilisateur clique sur un *Insight*, celui-ci est ajouté dans le filtre, et il en est retiré en cliquant dessus à nouveau. Les visualisations affichées sont alors celles qui correspondent à tous les *Insights* choisis.

5.4 Association entre données et variables visuelles

La Figure 7 présente la fenêtre permettant d'effectuer l'association des données aux variables visuelles de la visualisation sélectionnée. En collaboration avec des experts UX et après utilisation de *Catalogue Visu* par quelques utilisateurs, des fonctionnalités ont été ajoutées afin de guider ce choix.

Pour faciliter la compréhension des visualisations et aider l'utilisateur à associer ses données, une explication visuelle est disponible pour chaque variable des visualisations (Figure 9). De plus la réalisation de l'association entre données et variables visuelles s'adapte aux types de données. Certaines visualisations présentent des contraintes sur les données utilisées. Par exemple un graphique en courbes n'accepte pas de données qualitatives pour l'axe Y. Les données sont donc typées au moment de leur importation, et seuls les champs correspondant sont présentés à l'utilisateur.

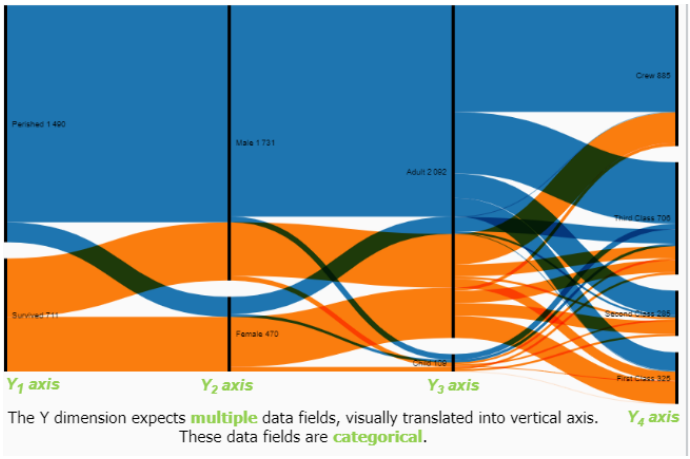

FIG. 9. Aide affichée pour effectuer l'association des données. Ici l'aide affichée est celle de la dimension Y des *Parallel Sets*.

Afin d'aider davantage l'utilisateur et d'éviter qu'il ne soit perdu, il est possible pour chaque visualisation de générer un exemple déjà paramétré. En cliquant sur le bouton dédié, des données d'exemple spécifiques à la visualisation sont chargées, l'association des données à la visualisation est réalisé automatiquement et la visualisation s'affiche. L'utilisateur peut alors modifier ces paramètres afin de mieux comprendre le fonctionnement de la visualisation, ou importer ses données afin de créer sa propre visualisation en s'inspirant de l'exemple.

5.5 Implémentation

Catalogue Visu est réalisée avec le framework *Angular*. Les visualisations sont réalisées à partir de la bibliothèque de visualisation *D3.js*. Elle se base sur un serveur *NodeJs* ainsi qu'une base de données *Neo4J* pour stocker les configurations des visualisations. L'application est actuellement à l'état de prototype, elle est en évolution permanente et de nouvelles versions sont régulièrement déployées. *Catalogue Visu* est accessible librement à l'adresse https://visual-catalog.forge.kaizen-solutions.net/.

6 DISCUSSION ET TRAVAUX FUTURS

Pour la version actuelle de notre prototype nous nous sommes concentrés sur un modèle binaire, que nous considérons comme une première étape avant l'évolution vers un autre modèle non binaire. Ce modèle n'est pas forcément adapté à une utilisation réelle, mais il nous a permis de simplifier les expérimentations sur la caractérisation des *Insights*. C'est pourquoi nous travaillons actuellement sur un modèle permettant d'exprimer un degré de satisfaction des *Insights* par les visualisations. Cette satisfaction peut être discrète ou continue. Nuancer le lien entre les *Insights* et les types de visualisation, par exemple "satisfait un peu, satisfait, satisfait complètement" serait un premier pas vers une représentation non binaire. Cette nuance laisserait plus de flexibilité à l'utilisateur pour exprimer son besoin de visualisation.

Les implémentations des visualisations que l'on peut trouver sur le web présentent de grandes variétés, en particulier au niveau des interactions. Ainsi un même type de visualisation peut satisfaire différemment les *Insights* en fonction des interactions proposées. Par exemple, un diagramme en bâtons permettant de sélectionner les bâtons pour avoir des informations facilitera la comparaison de données. Afin de prendre en compte cette diversité, nous prévoyons, en

plus de *D3.js*, l'intégration d'autres bibliothèques de visualisation. Aussi nous envisageons une étape supplémentaire de paramétrage de la visualisation. Deux types de paramétrages sont envisagés : le style, définissant par exemple les palettes de couleur, et les opérations possibles, qui définit les interactions réalisables. Ce dernier type de paramétrage est plus complexe, car il peut aller du simple choix d'opération effectuée lors d'un clic, jusqu'à la programmation d'opérations complexes comme du filtrage par drag&drop.

7 CONCLUSIONS

Dans la première partie de cet article, nous avons défini une taxonomie permettant aux utilisateurs de définir leurs besoins en visualisation afin de choisir des types de visualisation adéquats. Cette taxonomie ainsi que le choix de la visualisation ont été conçus et évalués itérativement par trois études utilisateur. Dans la seconde partie, nous avons mis en pratique cette méthode de choix avec *Catalogue Visu*, une application web de prototypage de visualisations. Le réalisation rapide de visualisations permet l'apprentissage de nouveaux types de visualisations, ainsi que le prototypages de visualisations développées pour des besoins spécifiques. Cette application est déployée et accessible librement, et en amélioration permanente. Par la suite nous prévoyons l'ajout de nouvelles fonctionnalités, comme le paramétrage par les utilisateurs des interactions disponibles dans les visualisations.

REMERCIEMENTS

Nous remercions particulièrement Franck Corduant et Olivier Desmaison pour leur appui et leur soutien à ce projet. Nous remercions Alexis Petit, Brieuc Horard, Julien Buisine, Rubal Ravinder, Noémie Rousseau, Morgane Delmas, Joachim Laviolette pour leur participation au projet. Enfin nous remercions tous nos autres collègues qui ont apporté des suggestions et de nouvelles idées pour la réalisation de cet article.

RÉFÉRENCES

[1] Hanene Azzag, Fabien Picarougne, Christiane Guinot, and Gilles Venturini. 2006. Vrminer : A tool for multimedia database mining with virtual reality. In *Processing and Managing Complex Data for Decision Support*. IGI Global, 318–339.

[2] Jacques Bertin and M Barbut. 1967. Sémiologie Graphique. Les diagrammes, les réseaux, les cartes.

[3] Fatma Bouali, Abdelheq Et-tahir Guettala, and Gilles Venturini. 2015. VizAssist : an interactive user assistant for visual data mining. *The Visual Computer* 32 (05 2015). https://doi.org/10.1007/s00371-015-1132-9

[4] Sylvain Bouveret, Renaud Blanch, Antoinette Baujard, François Durand, Herrade Igersheim, Jérôme Lang, Annick Laruelle, Jean-François Laslier, Isabelle Lebon, and Vincent Merlin. 2019. Voter Autrement 2017 - In Situ Experiment. (2019), 22 pages. https://doi.org/10.5281/zenodo.3548573

[5] Stuart Card. 2002. *Information Visualization.* L. Erlbaum Associates Inc., USA, 544–582.

[6] Stuart Card, Jock Mackinlay, and Ben Shneiderman. 1999. *Readings in Information Visualization : Using Vision To Think.* Journal Abbreviation : Information Visualization - IVS Publication Title : Information Visualization - IVS.

[7] Remco Chang, Tera Marie Green, William Ribarsky, and et al. 2009. Defining Insight for Visual Analytics.

[8] Yang Chen, Jing Yang, and William Ribarsky. 2009. Toward effective insight management in visual analytics systems. In *2009 IEEE Pacific Visualization Symposium*. IEEE, 49–56.

[9] Eun Kyoung Choe, Bongshin Lee, et al. 2015. Characterizing visualization insights from quantified selfers' personal data presentations. *IEEE computer graphics and applications* 35, 4 (2015), 28–37.

[10] Sarah Goodwin, Jason Dykes, Sara Jones, Iain Dillingham, Graham Dove, Alison Duffy, Alexander Kachkaev, Aidan Slingsby, and Jo Wood. 2013. Creative User-Centered Visualization Design for Energy Analysts and Modelers. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2516–2525. https://doi.org/10.1109/TVCG.2013.145

[11] Kyle Wm. Hall, Anthony Kouroupis, Anastasia Bezerianos, Danielle Albers Szafir, and Christopher Collins. 2022. Professional Differences : A Comparative Study of Visualization Task Performance and Spatial Ability Across Disciplines. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (Jan. 2022), 654–664. https://doi.org/10.1109/TVCG.2021.3114805

[12] Alaul Islam, Anastasia Bezerianos, Bongshin Lee, Tanja Blascheck, and Petra Isenberg. 2020. Visualizing Information on Watch Faces : A Survey with Smartwatch Users. In *2020 IEEE Visualization Conference (VIS)*. 156–160. https://doi.org/10.1109/VIS47514.2020.00038

[13] Pawandeep Kaur and Michael Owonibi. 2017. A review on visualization recommendation strategies. In *International Conference on Information Visualization Theory and Applications*, Vol. 4. SCITEPRESS, 266–273.

[14] Alicia Key, Bill Howe, Daniel Perry, and Cecilia Aragon. 2012. Vizdeck : self-organizing dashboards for visual analytics. In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*. 681–684.

[15] Ian Li, Anind K. Dey, and Jodi Forlizzi. 2012. Using Context to Reveal Factors That Affect Physical Activity. *ACM Trans. Comput.-Hum. Interact.* 19, 1, Article 7 (may 2012), 21 pages. https://doi.org/10.1145/2147783.2147790

[16] Zhengliang Liu, R. Crouser, and Alvitta Ottley. 2020. Survey on Individual Differences in Visualization. *Computer Graphics Forum* 39 (06 2020), 693–712. https://doi.org/10.1111/cgf.14033

[17] Ivan Logre and Anne-Marie Déry-Pinna. 2018. MDE in Support of Visualization Systems Design : a Multi-Staged Approach Tailored for Multiple Roles. *Proceedings of the ACM on Human-Computer Interaction* 2, EICS (2018), 1–17.

[18] Jock Mackinlay. 1986. Automating the design of graphical presentations of relational information. *Acm Transactions On Graphics (Tog)* 5, 2 (1986), 110–141.

[19] Eamonn James Maguire. 2014. *Systematising glyph design for visualization.* Ph.D. Dissertation. University of Oxford.

[20] Belgin Mutlu, Eduardo Veas, and Christoph Trattner. 2016. Vizrec : Recommending personalized visualizations. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 6, 4 (2016), 1–39.

[21] C. North. 2006. Toward measuring visualization insight. *IEEE Computer Graphics and Applications* 26, 3 (May 2006), 6–9. https://doi.org/10.1109/MCG.2006.70

[22] Catherine Plaisant, Jean-Daniel Fekete, and Georges Grinstein. 2008. Promoting Insight-Based Evaluation of Visualizations : From Contest to Benchmark Repository. *IEEE Transactions on Visualization and Computer Graphics* 14, 1 (Jan. 2008), 120–134. https://doi.org/10.1109/TVCG.2007.70412 Conference Name : IEEE Transactions on Visualization and Computer Graphics.

[23] Severino Ribecca. 2013. DataVizCatalog visual function taxonomy. https://datavizcatalogue.com/search.html. Accessed : 2020-09-29.

[24] P. Saraiya, C. North, and K. Duca. 2005. An insight-based methodology for evaluating bioinformatics visualizations. *IEEE Transactions on Visualization and Computer Graphics* 11, 4 (July 2005), 443–456. https://doi.org/10.1109/TVCG.2005.53 Conference Name : IEEE Transactions on Visualization and Computer Graphics.

[25] Hans-Jörg Schulz, Thomas Nocke, Magnus Heitzler, and Heidrun Schumann. 2013. A design space of visualization tasks. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2366–2375.

[26] Ben Shneiderman. 1996. The eyes have it : A task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE symposium on visual languages*. IEEE, 336–343.

[27] Melanie Tory and Torsten Moller. 2004. Rethinking visualization : A high-level taxonomy. In *IEEE Symposium on information visualization*. IEEE, 151–158.

[28] Andrea Vázquez-Ingelmo, Francisco J García-Peñalvo, and Roberto Therón. 2019. Capturing high-level requirements of information dashboards' components through meta-modeling. In *Proceedings of the Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality*. 815–821.

[29] Fernanda B Viegas, Martin Wattenberg, Frank Van Ham, Jesse Kriss, and Matt McKeon. 2007. Manyeyes : a site for visualization at internet scale. *IEEE transactions on visualization and computer graphics* 13, 6 (2007), 1121–1128.

[30] Kanit Wongsuphasawat, Dominik Moritz, Anushka Anand, Jock Mackinlay, Bill Howe, and Jeffrey Heer. 2015. Voyager : Exploratory analysis via faceted browsing of visualization recommendations. *IEEE transactions on visualization and computer graphics* 22, 1 (2015), 649–658.

[31] Angela Zoss. 2017. Data visualization : visualization types. *Duke Uni* (2017). https://guides.library.duke.edu/datavis/vis_types