



HAL
open science

Parallel Bubbles : Visualisation de données catégorielles dans des Coordonnées Parallèles

Raphaël Tuor, Florian Evéquo, Denis Lalanne

► **To cite this version:**

Raphaël Tuor, Florian Evéquo, Denis Lalanne. Parallel Bubbles : Visualisation de données catégorielles dans des Coordonnées Parallèles. Actes de la 28ième conférence francophone sur l'Interaction Homme-Machine, Oct 2016, Fribourg, Suisse. pp.299-306, 10.1145/3004107.3004142 . hal-01384343

HAL Id: hal-01384343

<https://hal.science/hal-01384343>

Submitted on 19 Oct 2016

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.

***Parallel Bubbles* : Visualisation de données catégorielles dans des Coordonnées Parallèles**

Raphaël Tuor

Centre de recherche Human-IST,
Faculté des sciences
1700, Fribourg, Suisse
raphael.tuor@unifr.ch

Florian Evéquoz

HES-SO Valais, IIG
3960, Sierre, Suisse
florian.evequoz@hes-so.ch

Denis Lalanne

Centre de recherche Human-IST,
Faculté des sciences
1700, Fribourg, Suisse
denis.lalanne@unifr.ch

Résumé

Dans cet article, nous discutons les techniques disponibles pour la représentation de données catégorielles au sein des Coordonnées Parallèles, une méthode de visualisation largement répandue dans les tâches d'analyse de données multivariées. Nous proposons les *Parallel Bubbles*, une méthode basée sur la fréquence améliorant la perception graphique des dimensions catégorielles dans les Coordonnées Parallèles. Nous réalisons une étude utilisateur comprenant des tâches de similarité et de fréquence, en comparant trois variations des Coordonnées Parallèles. Les *Parallel Bubbles* représentent un bon compromis en matière de performance pour les deux types de tâches et l'ajout d'un encodage visuel de fréquence provoque une différence significative dans la performance. Cette étude pose le premier jalon d'une série d'articles dont la suite aura comme objectif de tester les trois visualisations dans des tâches portant sur l'axe numérique, et dans lesquels le Parallel Set sera probablement moins performant.

Mots Clés

coordonnées parallèles ; données catégorielles ; étude

ACM Classification Keywords

H.5.2 Information interfaces and presentation [User Interfaces]

© ACM, 2016. This is the author's version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version was published in Actes de la 28ème conférence francophone sur l'Interaction Homme-Machine, 2016.
<http://dx.doi.org/10.1145/3004107.3004142>

Introduction

Les Coordonnées Parallèles [8] sont une méthode de visualisation populaire pour représenter des données multivariées. Les axes parallèles représentent les variables et chaque élément multivarié correspond à une polyligne traversant les axes. Cette méthode de visualisation permet de faire émerger des *patterns* rapidement [15], c'est donc un outil de choix dans les tâches d'analyse exploratoire. Un problème récurrent des Coordonnées Parallèles est l'enchevêtrement de lignes apparaissant lorsque le nombre de polygones est trop grand : il en devient difficile d'identifier des *clusters* et des tendances. La représentation de données discrètes augmente la confusion visuelle : les Coordonnées Parallèles « ne sont pas adaptées aux dimensions catégorielles : soit l'information de fréquence n'est pas visible, soit un classement est imposé à la représentation visuelle, influençant la perception des données » [10] (traduit). Les dimensions catégorielles sont représentées de la même manière que les dimensions continues : les Coordonnées Parallèles « implémentent un modèle de représentation continu, qui ne correspond pas au modèle discret de l'utilisateur » [10]. Un autre problème lié aux données catégorielles est l'*overplotting* : lorsque le nombre de catégories est significativement plus petit que le nombre d'objets, de nombreux objets partagent les mêmes catégories (voir Figure 1). Cela mène à « l'augmentation de la probabilité que de multiples lignes passent successivement par les mêmes points. Si plusieurs objets ont des valeurs identiques le long d'axes adjacents, une portion de leurs polygones se superposera exactement » [5]. Ces superpositions masquent l'information de fréquence et les *patterns*.

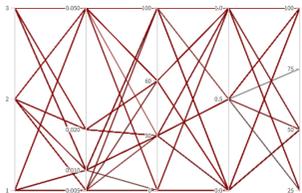


Figure 1: Exemple d'*overplotting* dans un graphique de Coordonnées Parallèles présentant des dimensions catégorielles ou discrètes.

Réduire l'*overplotting* et la confusion visuelle font partie des principaux défis dans la conception de graphiques de Coordonnées Parallèles [7], en particulier dans le cas de données catégorielles. Les chercheurs proposent plusieurs

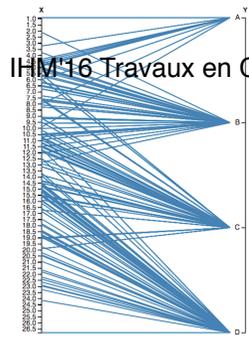
méthodes pour relever ces défis, mais « seul un nombre limité présente les résultats d'évaluations centrées sur l'utilisateur » [9] (traduit). Parmi celles-ci, les approches basées sur la fréquence sont prometteuses. Mais les bénéfices mesurables qui pourraient encourager l'utilisation de variantes « augmentées » des Coordonnées Parallèles font défaut. Il subsiste des doutes quant au potentiel de ces techniques d'aider les utilisateurs. Identifient-ils plus rapidement et précisément des *patterns* dans les données ? Interprètent-ils correctement les dimensions catégorielles ?

Dans cet article, nous donnons une vue d'ensemble des travaux réalisés sur le sujet. Nous présentons ensuite une variation des Coordonnées Parallèles, les *Parallel Bubbles*, censées améliorer la performance des utilisateurs dans des tâches de similarité et de fréquence. Pour mesurer cette performance, nous utilisons des données mixtes comprenant une dimension continue et une dimension catégorielle. Nous décrivons et discutons les résultats de cette expérience utilisateur dans la dernière section.

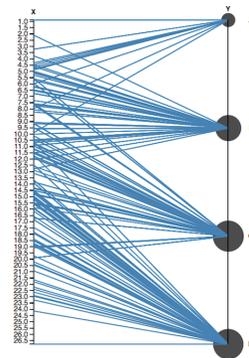
État de l'art

Les Coordonnées Parallèles [19] ont été initialement conçues pour représenter des dimensions continues. La divergence entre le modèle mental de l'utilisateur et la représentation des données peut être limitée, voire éliminée, avec l'utilisation de techniques basées sur la fréquence : les catégories sont représentées par des entités visuelles dont la taille correspond à leur fréquence [10]. Nous présentons ici les méthodes existantes.

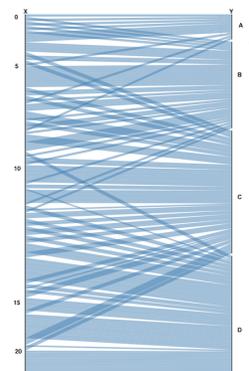
D'après Heinrich et Weiskopf [7], les approches de réduction de l'*overplotting* sont regroupées dans les techniques de filtrage, d'agrégation, de distorsion spatiale, de réarrangement des dimensions, et l'utilisation de couleurs.



(a) Coordonnées Parallèles



(b) Parallel Bubbles



(c) Parallel Sets

Figure 2: Les trois variantes de visualisation. L'axe gauche est continu, l'axe droit catégoriel. Le deuxième jeu de données (corrélation moyenne) est représenté.

Filtrage

Le filtrage consiste à enlever des signaux de l'entrée. Le *brushing* en est un bon exemple [14] : spécifier des plages de valeurs et utiliser des opérateurs logiques [11] permet de n'afficher qu'une portion des polygones, réduisant la confusion visuelle et l'*overplotting*. Cependant, le *brushing* standard est peu utile avec les axes catégoriels, l'ordre des catégories n'ayant pas de signification. Une manière d'implémenter le filtrage serait de permettre la sélection d'un nombre discret de catégories plutôt qu'une plage de valeurs.

Agrégation

Les techniques d'agrégation se basent sur le principe du regroupement des données et l'affichage d'objets agrégés plutôt que des objets individuels. De tels agrégats incluent la moyenne, la médiane ou le barycentre d'un sous-ensemble d'objets. Les fonctions de densité sont « particulièrement utiles pour révéler des aires et des groupes denses dans les données » [7] (traduit). Les représentations basées sur la fréquence montrent la distribution des points dans chaque catégorie. D'autres approches donnent une signification à l'espacement entre les axes catégoriels [13], utilisent des courbes de fréquence [3] ou encore construisent un intervalle proportionnel à la fréquence sur les axes catégoriels [18]. D'autres approches basées sur la géométrie lient les *clusters* à des enveloppes [12] ou à des *bounding boxes* [2].

Kosara *et al.* [10] ont proposé les *Parallel Sets*. Chaque catégorie est représentée par une boîte de taille proportionnelle à la fréquence de la catégorie (voir Figure 2c). Des parallélogrammes relient les axes catégoriels. Cette méthode est bien adaptée aux données purement catégorielles, mais ne semble pas idéale pour les jeux de données y mêlant des dimensions continues. Par défaut, chaque

valeur sur un axe de dimension continue devient une catégorie. De plus, il devient difficile de détecter les *outliers*. Les auteurs précisent que « montrer des axes continus comme de vraies dimensions de Coordonnées Parallèles serait évidemment l'affichage le plus adéquat pour ces données ».

Couleurs

L'utilisation de couleurs est « une méthode simple mais efficace pour distinguer un petit ensemble de catégories » [7] (traduit). Elle perd de son intérêt dès qu'il y a trop de catégories, ou plusieurs dimensions catégorielles concurrentes.

L'étude : Coordonnées Parallèles, Parallel Bubbles et Parallel Sets

Métaphore visuelle, hypothèse

Le but de notre étude est de comparer la performance de trois variations des Coordonnées Parallèles dans des tâches de similarité et de fréquence, afin de vérifier s'il existe une différence significative dans l'efficacité des participants. Notre approche, les *Parallel Bubbles*, doit permettre d'améliorer la visualisation des données catégorielles par rapport aux Coordonnées Parallèles standard en ajoutant un encodage visuel de la fréquence. La pierre angulaire des *Parallel Bubbles* consiste en une « bulle » (un rond) de rayon variable, qui représente une valeur catégorielle et dont la taille est proportionnelle au nombre d'objets qui possèdent cette catégorie. Le rayon de chaque bulle est défini comme la racine carrée du nombre d'objets, multipliée par deux. Des ensembles de bulles alignées verticalement sont disposés sur chaque axe de dimension catégorielle. Les dimensions continues sont représentées à la manière de Coordonnées Parallèles standard, offrant une distinction visuelle avec leurs homologues catégoriels. Chaque objet du jeu de données est représenté par une polygone passant par les axes continus et catégoriels. Cette approche donne une idée de la distribution des caté-

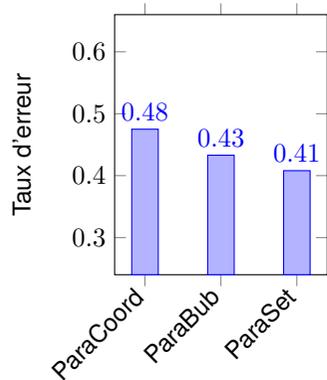


Figure 3: Taux d'erreur pour la tâche 1 (similarité) sur l'ensemble des données.

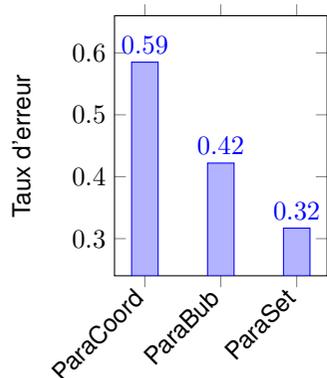


Figure 4: Taux d'erreur pour la tâche 3 (fréquence) sur l'ensemble des données.

gories au sein du jeu de données, et les « bulles » donnent une aide visuelle, restaurant l'information de fréquence perdue à cause de l'*overplotting* : la taille des bulles donne une information sur le nombre de lignes qui se recouvrent. L'hypothèse à tester est que les trois types de visualisations induisent une différence de performance dans les tâches de fréquence et de similarité. Nous formulons aussi les hypothèses suivantes : les Coordonnées Parallèles (ParaCoord) offrent les meilleures performances pour les tâches de similarité et les Parallel Sets (ParaSet) les meilleures performances pour les tâches de fréquence. Les *Parallel Bubbles* (ParaBub) représenteraient un compromis entre les trois méthodes. Les *Parallel Bubbles* offrent une meilleure performance que les Coordonnées Parallèles dans les tâches de fréquence, et de meilleures performances que les Parallel Sets dans les tâches de similarité. Nous avons utilisé trois variations de Coordonnées Parallèles (voir Figure 2) :

- **ParaCoord** – des Coordonnées Parallèles standard (Figure 2a). La visualisation est sensible aux problèmes de *clutter* et d'*overplotting*.
- **ParaBub** – des Coordonnées Parallèles avec ajout d'un encodage visuel de la fréquence des valeurs catégorielles (axe Y) sous forme de bulles de tailles variables (Figure 2b). La fréquence des valeurs catégorielles est plus aisément identifiable.
- **ParaSet** – une variation des Coordonnées Parallèles encodant toutes les fréquences d'après la hauteur du segment sur l'axe (Figure 2c). L'axe continu (X) est transformé en une dimension catégorielle. L'axe X est plus difficile à lire que dans les autres visualisations, car seules certaines valeurs y sont inscrites.

Afin d'évaluer la performance des participants sur les trois visualisations, nous avons soumis les participants à deux types de tâches, décrites dans la section suivante.

Tâches

Les tâches que nous avons évaluées ne représentent qu'un sous-ensemble des tâches couramment effectuées par les analystes de données, qui comprennent l'identification de clusters, d'outliers et le suivi d'individus, entre autres. D'après Fernstad et Johansson [1], « l'objectif général de l'analyse de données est d'identifier des structures et patterns dans les données. La plupart des patterns, comme les clusters et les corrélations, peuvent être définis en termes de similarité. Par conséquent, le type de tâche le plus intéressant est l'identification de relations en termes de similarité » (traduit). D'après ces auteurs, les tâches de fréquence présentent aussi un intérêt majeur. De manière analogue à [1], nous avons défini deux types de tâches :

1. **Similarité** – identifier des structures dans les données. Pour cela, pouvoir juger de la force de la corrélation entre deux dimensions est primordial.
2. **Fréquence** – estimer le nombre d'objets possédant une valeur catégorielle donnée. Identifier la catégorie la plus représentée est également important.

Sur cette base, voici les trois tâches qui ont été soumises :

- **(T1) Tâche de similarité** : Dans quelle mesure les deux dimensions sont-elles corrélées ? Réponses possibles : Pas du tout, modérément, beaucoup
- **(T2) Tâche de fréquence** : Quelle est la catégorie la plus représentée sur l'axe Y (catégoriel) ? Réponses possibles : A, B, C, D
- **(T3) Tâche de fréquence** : Quelle proportion du jeu de données possède la catégorie « C » (en %) ? Réponses possibles : de 0% à 100% par pas de 10%.

Les capacités d'interaction avec la visualisation ont été supprimées afin de minimiser le nombre de variables indépendantes et obtenir des résultats plus précis. On a ainsi rendu

impossibles le réarrangement manuel des axes, de la position des valeurs catégorielles sur les axes et la suppression, l'ajout et le regroupement de dimensions. Notre choix s'est porté sur des données à deux dimensions comportant une dimension continue X et une dimension catégorielle Y. Par construction, les graphiques de Coordonnées Parallèles permettent d'identifier facilement les relations entre dimensions voisines. Or, explorer des données de plus de deux dimensions au moyen des Coordonnées Parallèles revient à effectuer une succession de comparaisons de dimensions voisines. Les tâches que nous avons définies, portant sur des dimensions voisines, focalisent l'attention de l'utilisateur sur le plus petit niveau de granularité offert par les Coordonnées Parallèles.

Matériel d'étude

Les tâches ont été soumises sous la forme de questionnaires en ligne¹ et les participants recrutés sur la plateforme de crowdsourcing Prolific². 367 participants ont rempli le questionnaire. Aucun critère de sélection n'a été mis en place, un tutoriel garantissant la compréhension des visualisations. Chaque participant a été payé 0,70 £. Tester les méthodes de visualisation sur des jeux de données variés augmente la robustesse du résultat et il était nécessaire pour cela de contrôler le niveau de corrélation. Nous avons donc généré trois jeux de données, de taille suffisamment grande pour générer de l'*overplotting* (465 à 480 objets), à l'aide d'un script Python. La fonction `random.multivariate_normal(mean, cov, size)` de la librairie `numpy` nous a permis de générer des objets aléatoirement en suivant une distribution normale multivariée. Pour chaque jeu de données, nous avons exécuté cette méthode à trois reprises sur des plages de valeurs se superposant (ex. : [0, 20], [10, 30] [20, 40]), afin d'uniformiser

la distribution. L'argument `cov` contient la matrice de covariance de la distribution, qui permet de maîtriser la variation de la variable X par rapport à la variable Y en multipliant les éléments $C_{x,y}$ et $C_{y,x}$ par un indice prenant les valeurs 0,0 (pas de corrélation), 0,8 (corrélation moyenne) et 1,0 (forte corrélation) pour chaque jeu de données respectif.

Design expérimental

L'étude a pris la forme d'un « between-group » avec comme variable indépendante le type de visualisation (ParaCoord, ParaBub ou ParaSet). L'ordre de présentation des trois jeux de données a été contrebalancé en utilisant la procédure du carré Latin [4], donnant 6 variantes d'ordre des données par visualisation, pour un total de 18 questionnaires. Nous avons attribué un type de visualisation à chaque participant. Afin d'éviter tout biais lié à l'apprentissage, chacun des trois jeux de données n'a été montré qu'une fois à chaque participant, dans un ordre défini par le carré Latin. Chaque participant a suivi un tutoriel présentant la méthode de visualisation, puis a dû réaliser 3 tâches sur chacun des 3 jeux de données, pour un total de 9 tâches.

Procédure

Afin de niveler les capacités d'analyse visuelle des participants sur chaque méthode de visualisation, nous avons placé le tutoriel en début de questionnaire. Les participants ont été informés au début du questionnaire que ce seraient les performances des techniques de visualisation qui seraient évaluées, plutôt que leurs performances individuelles. Le tutoriel était suivi de la partie principale du questionnaire. À l'aide de la méthode de visualisation attribuée, il devait réaliser les trois tâches que nous avons décrites précédemment pour chaque jeu de données.

Résultats

Nous avons éliminé les résultats de 5 participants qui n'avaient rempli que la première tâche (1 pour ParaCoord, 4 pour Pa-

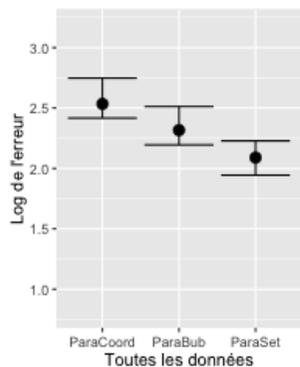


Figure 5: Moyenne des quartiles Q1 et Q3 et intervalle de confiance à 95% pour la tâche T2.

1. <https://www.surveymonkey.com/>

2. <https://www.prolific.ac/>

raSet). Nous avons analysé les réponses de 122 participants pour les Coordonnées Parallèles, 124 pour les *Parallel Bubbles* et 121 pour les Parallel Sets. Pour les tâches T1 (similarité) et T3 (fréquence), dont la réponse est dichotomique, nous avons calculé le taux d'erreur (voir Figure 3 et 4). Pour la tâche T2 (fréquence), la réponse était donnée par plage de 10%. Par souci de cohérence avec les autres études [6, 16], nous avons calculé le log de l'erreur absolue ainsi : $\log_2(|\text{valeurFournie} - \text{valeurCorrecte}| + \frac{1}{8})$. Une analyse de la variance (ANOVA) [4] sur les résultats des trois tâches a permis de rejeter l'hypothèse nulle avec comme significativité $p < 0,01$, confirmant l'hypothèse de l'influence du type de visualisation sur la performance. Pour la tâche T2, la moyenne des quartiles 1 et 3 pour le log de l'erreur absolue, ainsi que l'intervalle de confiance à 95% sont visibles dans le Tableau 1. Ces données sont présentées dans la Figure 5. Une décomposition par type de données (1 = non corrélées, 2 = moyennement corrélées, 3 = très corrélées) confirme les résultats globaux ; les données très corrélées provoquent le plus faible taux d'erreur (voir Figure 6). Les Coordonnées Parallèles sont systématiquement moins performantes que les deux autres méthodes. Cela confirme notre première hypothèse : l'ajout d'un encodage de la fréquence a une influence significative sur la performance des tâches de similarité et de fréquence. Pour la tâche T1 (similarité) et pour les trois jeux de données séparés, l'ANOVA n'a révélé une différence significative que pour le jeu de données 2 avec $p < 0,5$ et $F(3, 0205) = 5,7048$. Pour la tâche T2 (fréquence) et pour les trois jeux de données séparés, l'ANOVA a révélé une différence significative pour le jeu de données 1 avec $p < 0,5$ et $F(3, 0205) = 3,2678$ et le jeu de données 3 avec $p < 0,5$ et $F(3, 0205) = 4,7971$. Pour la tâche T3 (fréquence) et pour les trois jeux de données séparés, l'ANOVA a révélé une différence significative pour les trois jeux de données : pour les données 1 avec $p < 0,5$ et

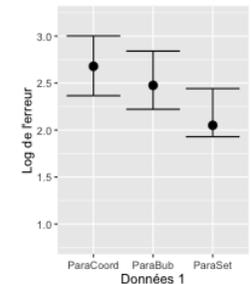
$F(3, 0205) = 5,029$, pour les données 2 avec $p < 0,001$ et $F(7, 0405) = 46,1558$ et pour les données 3 avec $p < 0,001$ et $F(7, 0405) = 47,7846$.

Type	Log erreur	IC 95%
ParaCoord	2,534	$\pm 0,165$
ParaBub	2,316	$\pm 0,159$
ParaSet	2,089	$\pm 0,142$

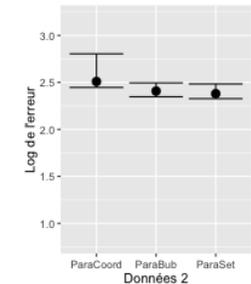
Tableau 1 : Moyenne des quartiles Q1 et Q3, et intervalles de confiance pour le log de l'erreur absolue, pour T2, par type de visualisation (ANOVA : $F(6, 9514) = 8,5469, p < 0,01$)

Conclusion et développements futurs

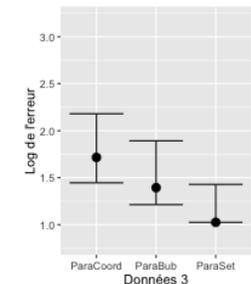
Notre hypothèse principale a été vérifiée : les trois types de visualisations induisent une différence de performance dans les tâches de fréquence et de similarité. Les résultats de l'évaluation montrent que les *Parallel Bubbles* représentent un compromis de performance entre les Parallel sets et les Coordonnées Parallèles. Ces dernières sont moins performantes dans toutes les tâches. Comme développements futurs, il serait intéressant d'étendre la méthode *Parallel Bubbles*, de l'implémenter dans un système fonctionnel, l'utiliser dans un contexte réel et la tester dans un cas d'utilisation. Une possibilité d'amélioration des *Parallel Bubbles* consisterait à définir le rayon des cercles comme proportionnel à la fréquence au lieu de la racine carrée utilisée ici, afin compenser les effets psychophysiques liés à la représentation sous forme d'aire [17]. On pourrait aussi examiner une alternative constituée de « sets » pour les axes catégoriels, combinés avec les Coordonnées Parallèles pour les axes continus. Les prochains articles compareront la performance des trois visualisations dans des tâches centrées sur l'axe continu, pour lesquelles les Parallel Sets seront probablement moins bons.



(a) Données non corrélées



(b) Données moyennement corrélées



(c) Données très corrélées

Figure 6 : Moyenne des centiles 20 et 80 et intervalle de confiance à 95% pour le log de l'erreur absolue pour T2, par type de données.

Bibliographie

- [1] Sara Johansson Fernstad and Jimmy Johansson. 2011. A Task Based Performance Evaluation of Visualization Approaches for Categorical Data Analysis. *2011 15th International Conference on Information Visualisation* (jul 2011), 80–89. DOI : <http://dx.doi.org/10.1109/IV.2011.92>
- [2] Ying-Huey Fua Ying-Huey Fua, M.O. Ward, and E.a. Rundensteiner. 1999. Hierarchical parallel coordinates for exploration of large datasets. *Proceedings Visualization '99 (Cat. No.99CB37067)* (1999), 43–508. DOI : <http://dx.doi.org/10.1109/VISUAL.1999.809866>
- [3] Zhao Geng, Zhenmin Peng, Robert S. Laramee, Rick Walker, and Jonathan C. Roberts. 2011. Angular histograms : Frequency-based visualizations for large, high dimensional data. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 2572–2580. DOI : <http://dx.doi.org/10.1109/TVCG.2011.166>
- [4] Anthony M. Graziano and Michael L. Raulin. 2010. *Research Methods - A Process of Inquiry*. 400 pages.
- [5] Susan L Havre, Anuj Shah, Christian Posse, and Bobbie-Jo Webb-Robertson. 2006. Diverse information integration and visualization, Vol. 6060. 60600M–60600M–11. DOI : <http://dx.doi.org/10.1117/12.643492>
- [6] Jeffrey Heer and Michael Bostock. 2010. Crowdsourcing Graphical Perception : Using Mechanical Turk to Assess Visualization Design. *Proceedings of the 28th Annual Chi Conference on Human Factors in Computing Systems* (2010), 203–212. DOI : <http://dx.doi.org/10.1145/1753326.1753357>
- [7] J Heinrich and D Weiskopf. 2013. State of the Art of Parallel Coordinates. *Eurographics* (2013), 95–116. DOI : <http://dx.doi.org/10.2312/conf/EG2013/stars/095-116>
- [8] A Inselberg and Bernard Dimsdale. 1990. Parallel coordinates : a tool for visualizing multi-dimensional geometry. (1990). DOI : <http://dx.doi.org/10.1109/VISUAL.1990.146402>
- [9] Jimmy Johansson and Camilla Forsell. 2016. Evaluation of Parallel Coordinates : Overview, Categorization and Guidelines for Future Research. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 579–588. DOI : <http://dx.doi.org/10.1109/TVCG.2015.2466992>
- [10] Robert Kosara, Fabian Bendix, and Helwig Hauser. 2006. Parallel sets : Interactive exploration and visual analysis of categorical data. In *IEEE Transactions on Visualization and Computer Graphics*, Vol. 12. IEEE, 558–568. DOI : <http://dx.doi.org/10.1109/TVCG.2006.76>
- [11] A.R. Martin and M.O. Ward. 1995. High Dimensional Brushing for Interactive Exploration of Multivariate Data. *Proceedings Visualization '95* (1995), 271–278. DOI : <http://dx.doi.org/10.1109/VISUAL.1995.485139>
- [12] Rida E. Moustafa. 2011. Parallel coordinate and parallel coordinate density plots. *Wiley Interdisciplinary Reviews : Computational Statistics* 3, 2 (2011), 134–148. DOI : <http://dx.doi.org/10.1002/wics.145>
- [13] Geraldine E. Rosario, Elke A. Rundensteiner, David C. Brown, and Matthew O. Ward. 2003. Mapping nominal values to numbers for effective visualization. In *Proceedings - IEEE Symposium on Information Visualization, INFO VIS*. 113–120. DOI : <http://dx.doi.org/10.1109/INFVIS.2003.1249016>
- [14] Ben Shneiderman. 1994. Dynamic queries for visual information seeking. *IEEE Software* 11, 6 (1994), 70–77. DOI : <http://dx.doi.org/10.1109/52.329404>
- [15] Harri Siirtola, Tuuli Laivo, Tomi Heimonen, and Kari-Jouko Rähkä. 2009. Visual Perception of Parallel Coordinate Visualizations. In *2009 13th International Conference Information Visualisation*. IEEE, 3–9. DOI : <http://dx.doi.org/10.1109/IV.2009.25>

- [16] Drew Skau and Robert Kosara. 2016. Arcs , Angles , or Areas : Individual Data Encodings in Pie and Donut Charts. 35, 3 (2016). DOI : <http://dx.doi.org/DOI:10.1111/cgf.12888>
- [17] SS Stevens. 1975. Psychophysics : Introduction to its perceptual, neural and social prospects. (1975).
- [18] Soon Tee Teoh and Kwan-Liu Ma. 2003. Painting-Class : interactive construction, visualization and exploration of decision trees. *Star* (2003), 667–672. DOI : <http://dx.doi.org/10.1145/956750.956837>
- [19] Edward J. Wegman. 1990. Hyperdimensional Data Analysis Using Parallel Coordinates. *J. Amer. Statist. Assoc.* 85, 411 (sep 1990), 664–675. DOI : <http://dx.doi.org/10.1080/01621459.1990.10474926>