

## Évaluation des interfaces visuelles

Guy Melançon, Monique Noirhomme-Fraiture, Bruno Pinaud

► **To cite this version:**

Guy Melançon, Monique Noirhomme-Fraiture, Bruno Pinaud. Évaluation des interfaces visuelles. Atelier Visualisation d'informations, interaction et fouille de données - Conférence Extraction et Gestion des Connaissances 2013 (EGC'2013), Jan 2013, Toulouse, France. <hal-00772804>

**HAL Id: hal-00772804**

**<https://hal.archives-ouvertes.fr/hal-00772804>**

Submitted on 11 Jan 2013

**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.

# Évaluation des interfaces visuelles

Guy Melançon\*, Monique Noirhomme-Fraiture\*\* et Bruno Pinaud\*

\*Université de Bordeaux, UMR CNRS 5800 LaBRI  
351 Cours de la Libération, 33405 Talence Cedex, France  
{guy.melancon, bruno.pinaud}@labri.fr,  
<http://www.labri.fr>

\*\*Université de Namur, Faculté d'Informatique  
21 rue Grandgagnage, 5000 Namur, Belgique  
mno@info.fundp.ac.be  
<http://www.info.fundp.ac.be>

**Résumé.** La présente contribution est une synthèse de l'atelier VIF qui a eu lieu le 16 octobre 2012 à Bidart (France, 64), dans le cadre de la conférence Ergo IHM 2012. Le thème de l'atelier était l'évaluation des interfaces visuelles en analyse de données sous la forme d'une table ronde. Pendant l'atelier, les aspects suivants du contraignant, long et complexe processus d'évaluation ont été abordés : précisions de quelques concepts ; état des lieux des modes d'évaluation des interfaces dans les communautés IHM et analyse de données ; bonnes pratiques d'expérimentation utilisateurs ; problèmes de l'évaluation de systèmes complexes. Les organisateurs, encouragés par les résultats de cet atelier et les discussions avec les participants ont souhaité rédiger un document pour le partager avec un plus grand nombre de chercheurs.

## 1 Introduction

La présente contribution est une synthèse de l'atelier VIF<sup>1</sup> qui a eu lieu le 16 octobre 2012 à Bidart (France, 64), dans le cadre de la conférence Ergo IHM 2012. Le thème de l'atelier était l'évaluation des interfaces visuelles en analyse de données sous la forme d'une table ronde. Les organisateurs, encouragés par les résultats de cet atelier et les discussions avec les participants ont souhaité rédiger un document pour le partager avec un plus grand nombre de chercheurs.

Pendant l'atelier, les présentations et la discussion ont porté sur les aspects suivants : précisions de quelques concepts du contraignant, long et complexe processus d'évaluation ; état des lieux des modes d'évaluation des interfaces dans les communautés IHM et analyse de données ; bonnes pratiques d'expérimentation utilisateurs ; problèmes de l'évaluation de systèmes complexes. Nous suivrons ce plan pour le présent document.

Les constatations reprises dans les paragraphes 3 et 4 proviennent principalement de notre expérience comme relecteur pour des conférences et journaux, en particulier EGC, IHM, INTERACT, VisWeek (et plus généralement IEEE TVCG), EuroVis, PacificVis et Graph Drawing

---

1. Groupe de travail commun des associations EGC et AFIHM sur la visualisation d'informations, l'interaction et la fouille de données.

(GD). Il ne nous est donc pas possible de fournir des références précises. Néanmoins, l'évaluation devient un sujet de plus en plus présent et l'ouvrage récent d'Helen Purchase (Purchase, 2012) témoigne de cette prise de conscience de la communauté et des efforts de structuration engagés notamment depuis (Plaisant, 2004) et (Thomas et Cook, 2005).

## 2 Différentes méthodes d'évaluation

Nous nous concentrons ici sur des méthodes qui incluent des utilisateurs finaux, par opposition aux méthodes à base d'experts. En effet, en IHM, il est fréquent que l'on fasse appel au concours d'un ou plusieurs experts en ergonomie. Ceux-ci utilisent le logiciel en se mettant à la place d'un utilisateur et décèlent les défauts de l'interface. Leur bonne connaissance des règles et principes d'ergonomie ainsi que l'usage d'heuristiques leur permettent de mettre en évidence des défauts majeurs. Cependant, même en faisant appel à plusieurs experts, on ne peut déceler qu'une partie des problèmes rencontrés par les utilisateurs. Il a été montré qu'il n'était pas intéressant d'utiliser plus de cinq experts car au-delà de ce nombre, les nouveaux experts n'identifient plus de nouveaux défauts (Nielsen et Molich, 1990). Ajoutons qu'il est aussi possible de tester automatiquement les interfaces sur base de règles. De nombreux systèmes ont été créés pour évaluer l'accessibilité de pages Web comme par exemple DESTINE (Beirekdar et al., 2002, 2005). Comme ces systèmes utilisent le code HTML des pages, ils ne couvrent que 70% des règles. Toutes les règles de type sémantique sont alors ignorées.

Les méthodes basées sur des utilisateurs peuvent être réparties en trois classes : les études de cas, les enquêtes et les expérimentations.

**Les études de cas** consistent à décrire en détail l'usage de l'interface par quelques utilisateurs finaux bien choisis. Sans qu'on puisse réellement généraliser les informations obtenues à l'ensemble de la population, les études peuvent révéler des difficultés et des comportements d'utilisation.

**Les enquêtes** à plus grande échelle sont maintenant facilitées par l'usage d'Internet (comme c'est le cas aux USA avec le site "Mechanical Turk" d'Amazon<sup>2</sup>, par exemple). Des participants, choisis à l'avance ou volontaires, sont invités à se connecter sur un site et à tester un logiciel en ligne. Habituellement, un scénario leur est proposé. Après le test, ils remplissent un questionnaire portant, d'une part, sur des réponses correspondant au scénario et, d'autre part, sur leurs impressions personnelles. Des questions permettant d'identifier leur profil sont également ajoutées (sexe, âge, profession, formation, ...). Cette méthode est assez rapide à mettre en place car elle ne demande pas d'observation directe. Elle permet d'avoir plus facilement un grand échantillon. Toutefois, beaucoup de biais peuvent se glisser dans l'étude étant donné le manque de contrôle (surtout par Internet). En particulier, on risque des réponses peu fiables et un biais dans le profil des répondants. Des corrections sont toutefois possibles. Lorsqu'il apparaît que le questionnaire a été rempli "au hasard", l'enquêteur a le loisir de l'écarter. Comme dans une enquête par téléphone, il peut aussi rétablir un équilibre dans le profil des participants en relançant l'enquête auprès des profils manquants. Enfin, on peut espérer qu'un grand nombre de réponses lisse les questionnaires erronés.

Étant donné qu'en IHM et en analyse des données, la plupart des études portent sur l'expérimentation, nous nous concentrerons dans la suite sur cette technique.

---

2. <https://www.mturk.com/mturk/welcome>

### 3 État des lieux en visualisation des données

Dans la suite, nous entendons par “visualisation des données” également “visual analytics” qui est une méthode de plus en plus populaire d’analyse utilisant le visuel comme support. Il ne faut pas confondre avec la visualisation en analyse de données qui est utilisée pour visualiser le résultat de l’analyse. L’IHM est alors l’interface/interactions entre homme et machine.

Dans ce domaine, on rencontre très peu d’expérimentations avec des utilisateurs et bien sûr pas d’enquête de grande envergure.

Suivant la pratique des chercheurs en analyse de données, les méthodes sont bien validées sur plusieurs jeux de données (simulées ou réelles). Les résultats sont comparés avec des méthodes connues. Mais toutes les manipulations sont effectuées par les auteurs des méthodes. Il n’y a donc aucun retour de la part des utilisateurs potentiels. Cette étape est cependant essentielle. Ce n’est pas parce qu’une méthode performante permet à son auteur de découvrir des propriétés intéressantes dans les données qu’elle convient à tout le monde. Si l’utilisateur potentiel n’est pas capable de s’en servir, la méthode ne sera jamais utilisée. C’est ce qui est arrivé à bon nombre de logiciels, abandonnés parce que leur interface n’était pas adaptée au profil des utilisateurs.

### 4 État des lieux en IHM

Il y a une douzaine d’années, les publications scientifiques ne relataient pas ou très peu d’expérimentations. Tout au plus, les nouvelles interfaces étaient-elles testées de manière informelle avec quelques cobayes, proches du développeur. Mais principalement depuis 2004 grâce aux travaux de C. Plaisant (Plaisant, 2004), la communauté internationale de visualisation d’informations a commencé à traiter le problème de l’évaluation en proposant de hiérarchiser les différents types d’évaluation possibles en fonction des objectifs recherchés (cf. Thomas et Cook, 2005, chap. 6). L’objectif général des chercheurs de cette communauté est de permettre à des utilisateurs experts de leurs données de les manipuler par des métaphores visuelles. On comprend donc, l’intérêt de vérifier précisément que les solutions fournies répondent aux problèmes des utilisateurs. Néanmoins, il n’existe aucune méthode générique d’évaluation. Un article récent, résultat d’un travail gigantesque d’analyse de 850 publications dans le domaine de la visualisation d’informations recense 7 scénarios possibles d’évaluation (Lam *et al.*, 2012). Chaque scénario définit un objectif précis pour l’évaluation et la méthode pour y parvenir.

De nos jours, grâce à ces travaux et sous la pression des relecteurs des conférences internationales, la situation s’est améliorée : des expérimentations sont soigneusement décrites, mais souvent avec un petit nombre de participants. Les hypothèses sont bien posées, le processus est reproductible. Néanmoins, on rencontre très souvent le terme “évaluation” mais ce dernier nous semble sur-estimé. Le protocole employé ressemble plus à “une validation” de la solution proposée. De notre point de vue, une évaluation doit faire l’objet d’un article entier alors qu’une validation est un processus plus rapide à effectuer et peut donc n’être qu’une partie d’un article. Une validation consiste bien souvent à prendre 4 à 5 cobayes qui effectuent quelques tâches simples pour montrer que la solution fonctionne. Cette méthode ne permet évidemment pas d’effectuer une analyse statistique des résultats (bien que certains auteurs s’y essaient). Plus globalement, du côté de l’analyse statistique, la situation reste malheureusement critiquable :

## Évaluation des interfaces visuelles

- dans de nombreux cas, les auteurs se contentent de publier des boîtes à moustaches, sans test et donc sans possibilité d’inférence ;
- l’échantillon, peu nombreux, est fortement biaisé car constitué, en général, de jeunes informaticiens âgés de 20 à 35 ans ;
- si un test statistique est utilisé, il s’agit généralement de l’ANOVA car ce test est à la mode dans certaines communautés. Il n’y a bien souvent pas de vérification des indispensables hypothèses de normalité, ni d’égalité des variances. L’ANOVA est parfois même utilisée pour une comparaison entre deux moyennes de données appariées alors que le test  $t$  dit de “student” (équivalent à l’ANOVA dans ce cas) est plus simple à mettre en place. Il semble que de nombreux chercheurs se contentent d’une analyse automatique type “presse bouton”, sans bien comprendre la méthode utilisée ;
- Lorsque les hypothèses de normalité et d’égalité des variances ne sont pas satisfaites, les tests non paramétriques (test de Wilcoxon par exemple), applicables avec peu de données, sont le plus souvent ignorés ;
- il arrive même que le test utilisé soit erroné. Par exemple, prenons une expérimentation qui consiste à tester différentes interfaces et manipulations par les mêmes cobayes. Il faut impérativement utiliser une variante de l’ANOVA dite “within” ou “mixte” ou “inter-participants”. Autre exemple, la correction de Bonferroni nécessaire lors de l’analyse répétée d’un même jeu de données est très rarement employée. Ces problèmes de méthodologie peuvent conduire à des résultats complètement erronés qui peuvent alors remettre en question les conclusions de l’expérimentation.

Notons que pour un relecteur, il n’est pas aisé d’identifier les erreurs car les articles fournissent bien souvent peu d’information : ils ne citent pas le test utilisé (et parfois même pas le seuil de décision  $\alpha$ ) et se contentent de donner la “ $p$ -value”. Lors de l’utilisation d’une ANOVA, ils donnent une valeur de  $F$  sans donner les nombres de degrés de liberté (*idem* pour le  $t$  de student), *etc.* Parfois, un  $p$  anormalement petit, vu le nombre d’observations, peut laisser suspecter une erreur grossière mais le relecteur n’a aucune certitude sur ce point. D’autre part, certains relecteurs, incompetents en statistiques, n’identifient pas les erreurs ou, *a contrario*, font des remarques tout à fait incorrectes ou incohérentes.

## 5 Bonnes pratiques en évaluation

La conduite d’une évaluation est une tâche difficile et chronophage sachant qu’il n’est souvent pas possible (par manque de temps) de tester toutes les combinaisons possibles des différents paramètres. L’évaluation doit donc suivre un protocole précis et rigoureux afin d’éviter d’être mise en doute. Tous les choix et les nécessaires compromis effectués doivent être clairement justifiés. L’expérimentation doit aussi être juste. Elle ne doit pas favoriser une méthode par rapport à une autre sur n’importe quel aspect (bonne/mauvaise implémentation, tâche orientée vers une méthode plutôt que l’autre, ...). Il faut être conscient que la généralisation des résultats finaux ne peut de toute façon qu’être limitée aux conditions de l’expérimentation. Les différentes bonnes pratiques énoncées dans cette partie sont issues de travaux d’évaluation (fructueux ou non) menés en collaboration avec Helen Purchase de l’université de Glasgow. Elle a récemment publié un ouvrage détaillé sur ce sujet (Purchase, 2012). Les paragraphes suivants ne sont en aucun cas exhaustif (se référer à l’ouvrage de H. Purchase pour plus de détails). Nous citons ci-dessous les principaux points stratégiques pour la bonne conduite d’une

expérimentation qui permettent de prévenir les problèmes et donc limiter les biais qui auront une influence néfaste sur la qualité des données recueillies.

En amont de l'expérimentation, il est impératif d'identifier clairement les objectifs en les formulant sous la forme d'une ou plusieurs questions de recherche et comment y répondre. Ces éléments vont ensuite servir à formuler les tâches que les participants auront à résoudre et les mesures à effectuer (bien souvent le taux de mauvaise réponse et le temps de réponse). Au sujet de la difficulté des tâches, elles doivent être discriminantes entre les méthodes tout en évitant les effets planchers et plafonds : si une tâche est trop facile, on n'observe aucune différence entre les méthodes (effet plancher) et à l'inverse, si une tâche est trop difficile (effet plafond), le participant se décourage et n'aboutit pas (les réponses sont alors quasiment toutes mauvaises). Pour éviter ces phénomènes désagréables, il convient de toujours tester l'expérimentation sur un petit nombre de cobayes afin de la calibrer : difficulté (raisonnable) des tâches, rédaction du questionnaire final (validation), durée (raisonnable) de l'expérimentation, *etc.* Ces tests sont appelés des expériences pilotes. Une fois que le protocole d'expérimentation est clairement défini, il ne doit plus être modifié. Chaque participant doit subir exactement la même expérimentation sous peine de devoir supprimer toutes les données recueillies et recommencer l'expérimentation depuis le début (en trouvant de nouveaux participants).

Pendant l'expérimentation, il est impératif de s'assurer en permanence de la qualité des données recueillies. Par exemple, les participants doivent tous passer l'expérimentation dans les mêmes conditions expérimentales, l'expérimentateur doit s'assurer que le cobaye répond en toute bonne foi aux questions, la plate-forme d'évaluation doit être robuste, *etc.* L'expérimentateur doit donc rester avec le participant jusqu'à la fin de l'expérimentation (et être disponible à tout moment si besoin) pour s'assurer que le processus d'évaluation se déroule correctement. Pour conserver la pleine et entière collaboration des participants, il faut aussi prendre en compte leur inévitable fatigue. L'expérimentation doit donc comporter des temps de repos régulièrement positionnés. La fatigue peut aussi être minimisée en regroupant les tâches équivalentes pour minimiser les changements de contexte cognitif. La charge cognitive des participants doit aussi être prise en compte. Ils doivent se concentrer au maximum sur la tâche à résoudre plutôt que de perdre du temps sur le fonctionnement de l'interface d'évaluation (qui doit donc être la plus simple possible) ou la réponse aux questions (pas de texte libre notamment). De plus, si les participants sont motivés par une récompense, les résultats sont bien souvent de meilleure qualité.

Après l'expérimentation, il est souhaitable de recueillir des informations qualitative de la part des participants et leurs impressions sur l'ensemble du processus. Ces nouvelles données vont servir à pondérer les résultats récupérés et affiner leur analyse.

## 6 Problèmes de l'évaluation de systèmes complexes

La nature des expériences qui sont menées, et l'analyse statistique qui est ensuite effectuée sur ses résultats, exige de pouvoir "mesurer" les temps d'accomplissement des tâches et le taux de réussite (ou plus souvent le nombre d'erreurs) de l'utilisateur. Cela conduit souvent à définir des tâches de bas niveau, en terme de manipulation des données. Or, l'objectif ultime de la visualisation analytique est de produire des interfaces capables de seconder l'utilisateur dans la formulation d'hypothèses sur le problème étudié, voire dans la découverte de nouvelles connaissances (van Wijk, 2005). Cet objectif est difficilement compatible avec les exigences

## Évaluation des interfaces visuelles

des expériences contrôlées. Munzner (2009) propose un modèle explicitant la conception et la validation des systèmes ou des techniques de visualisation sur quatre couches conceptuelles imbriquées (Figure 1). Il est utile de le décrire ici afin de bien cerner la portée des expérimentations contrôlées lorsque vient le moment d'affirmer l'efficacité et l'utilisabilité d'une technique ou d'un algorithme.

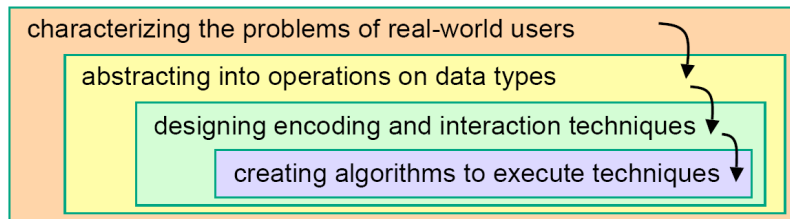


FIG. 1 – Le modèle en couches imbriquées de Munzner (Munzner, 2009).

La couche de plus haut niveau exige une bonne définition du domaine de référence, des questions qu'il s'agit d'étudier et des réponses attendues. Dans un contexte de visualisation de données bio-informatique, il pourrait s'agir de questions relatives à l'identification de familles de gènes comme bio-marqueurs d'une pathologie, par exemple. Dans un contexte de supervision d'un réseau informatique hautement sécurisé, on pourrait chercher à identifier des utilisateurs faisant transiter du contenu de manière illégale (en le communiquant à des tiers non autorisés), par exemple<sup>3</sup>.

De cette couche doit pouvoir émerger la couche inférieure donnant lieu à la spécification de tâches à effectuer sur un jeu de données. Il ne s'agit pas encore ici de tâches décrites en termes de manipulation d'une interface (qui reste à spécifier et concevoir), mais bien de tâches d'analyse à conduire sur un jeu de données lui-même spécifié par les questions formulées au niveau le plus haut. Dans les exemples donnés au paragraphe précédent, ces tâches deviennent par exemple "Identifier les familles de gènes dont les niveaux d'expression sont très similaires dans un certains nombres de conditions expérimentales" ou "Identifier les acteurs dont l'usage du réseau paraît anormal, tant au niveau des débits, des destinations que des plages horaires où ces irrégularités se produisent".

C'est ici que le concepteur doit traduire ces spécifications en termes de manipulations au niveau de l'interface. Cela exige aussi de préciser les représentations visuelles et les indices visuels qui seront adoptés. Les tâches prennent à ce (troisième) niveau la forme d'une série de manipulation qu'il est possible de faire pour répondre aux questions posées. C'est ici que l'évaluation par expériences contrôlées intervient, puisqu'il est nécessaire de s'assurer de proposer à l'utilisateur la/les représentation/s et les interactions les plus performantes pour les tâches qu'il a à accomplir.

Munzner ajoute encore un quatrième niveau où entre en jeu la performance des algorithmes. En effet, il est primordial d'implémenter des algorithmes performants, tant en temps qu'en justesse des résultats obtenus si on veut espérer proposer une interface de visualisation efficace.

3. Cet exemple s'inspire du concours IEEE VAST 2009. Voir <http://hcil.cs.umd.edu/localphp/hcil/vast/index.php>.

Le modèle de Munzner est clair sur un point qui nous intéresse ici : la performance seule d'un algorithme n'assure en rien de l'efficacité de la technique qui l'exploite. De même, il met en garde contre toute affirmation du caractère universelle d'une technique en vertu de sa supériorité démontrée dans le cadre d'une expérience contrôlée. L'utilisabilité tient aussi au contexte, du domaine d'application, des données qui forment le support de l'exploration et/ou de l'analyse et des manipulations –des tâches– qu'il convient d'y mener. Une technique performante pour certaines tâches pourraient bien s'avérer inutile ou inefficace dans un contexte différent.

## 7 Défis majeurs et conclusion

L'évaluation des interfaces et interactions est un problème difficile qui n'admet pas de solution universelle. Une expérimentation doit suivre un protocole particulier afin que les résultats obtenus puissent être analysés. Pour aller plus loin, la conception des outils doit-elle aussi répondre à un protocole précis afin de bien cerner quelles parties concernent l'utilisateur et donc les points à évaluer. Nous relevons quelques défis majeurs pour améliorer les bonnes pratiques pour l'expérimentation des interfaces. Les problèmes cités ci-dessous peuvent en partie être résolu par l'organisation de séances de tutorats pour expliquer les bonnes pratiques en évaluation :

**Améliorer la qualité des échantillons :** taille suffisante, échantillon représentatif de la population cible (personnes âgées, tous genres confondus, jeunes, ...);

**En visualisation de données, mieux identifier l'utilisateur :** Est-ce un statisticien ou le propriétaire des données ? S'il s'agit de cette dernière catégorie, quelle formation a-t-il ou quel effort est-il prêt à faire pour utiliser le logiciel ?;

**Mieux sensibiliser les chercheurs aux évaluations :** si un chercheur met au point une solution inutilisable et incompréhensible pour l'utilisateur spécialiste des données, le résultat est un temps précieux gâché pour tout le monde ;

**Améliorer la formation des informaticiens en statistique :** l'analyse des résultats et la présence (ou pas) de différence significative ne peut être validé que par des tests statistiques correctement effectués pour montrer que les différences sont significatives (ou pas) ;

Pour aller plus loin, les évaluations standards de type essais-erreurs (comptage des mauvaises réponses et du temps de réponse) commencent à montrer leurs limites. La communauté travaille à la mise au point de nouvelles méthodes d'évaluation. Le colloque bi-annuel BELIV<sup>4</sup> est une première réponse pour faire émerger de nouvelles techniques. On peut notamment citer l'utilisation d'un oculomètre (eye-tracker). En effet, le cobaye n'exploite peut-être pas l'interface en cours d'évaluation comme l'avait pensé son concepteur. Une étude des données du regard de l'utilisateur semble donc s'avérer fort judicieuse. Néanmoins, les oculomètres délivrent un flot de données extrêmement important. Il faut donc avoir recours à des techniques d'analyse et de fouilles de données pour exploiter ces résultats.

## Remerciements

Ce travail a été en partie financé par le projet ANR EVIDEN ANR 2010-JCJC-0201-01.

---

4. Beyond Time and Errors : Novel Evaluation Method for Visualization (BELIV), <http://www.beliv.org>



## Références

- Beirekdar, A., M. Keita, M. Noirhomme, F. Randolet, J. Vanderdonckt, et C. Mariage (2005). Flexible reporting for automated usability and accessibility evaluation of web sites. In *Proc. of the 2005 IFIP TC13 Int. Conf. on Human-Computer Interaction*, INTERACT'05, pp. 281–294.
- Beirekdar, A., J. Vanderdonckt, et M. Noirhomme-Fraiture (2002). A framework and a language for usability automatic evaluation of web sites by static analysis of html source code. In *Proc. of the 4<sup>th</sup> Int. Conf. on Computer-Aided Design of User Interfaces, CADUI'02*, pp. 337–348.
- Lam, H., E. Bertini, P. Isenberg, C. Plaisant, et S. Carpendale (2012). Empirical studies in information visualization : Seven scenarios. *IEEE Trans. on Visualization and Computer Graphics* 18(9), 1520–1536.
- Munzner, T. (2009). A nested process model for visualization design and validation. *IEEE Trans. on Visualization and Computer Graphics* 15, 921–928.
- Nielsen, J. et R. Molich (1990). Heuristic evaluation of user interfaces. In *CHI'90. Proc. of the SIGCHI Conf. on Human Factors in Computing*, pp. 249–256.
- Plaisant, C. (2004). The challenge of information visualization evaluation. In *Proc. of the Working Conf. on Advanced Visual Interfaces, AVI '04*, pp. 109–116. ACM.
- Purchase, H. C. (2012). *Experimental Human-Computer Interaction. A practical Guide with Visual Examples*. Cambridge University Press.
- Thomas, J. J. et K. A. Cook (Eds.) (2005). *Illuminating the Path : The Research and Development Agenda for Visual Analytics*. IEEE Computer Society.
- van Wijk, J. J. (2005). The value of visualization. In *Proc. of IEEE Visualization*, pp. 79–86.

## Summary

This contribution is a synthesis of the VIF panel discussion of the Ergo IHM 2012 conference (October 16<sup>th</sup>, 2012 in Bidart, France). The panel covered the evaluation of visual interface in data analysis. The following themes of the long and difficult evaluation process have been addressed: precisions of some concepts, state of the art of the different evaluation modes in HCI and Data Analysis; good practices for user evaluations; complex systems evaluations. The organizers wrote this document to share the interesting results of the panel with a broader range of researchers.