TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

CRIM   Centre de recherche informatique de Montréal

ETS   École de technologie supérieure

HSM   Horizontal small multiples

MWR   Money-weighted rate of return

TWR   Time-weighted rate of return

VSM   Vertical small multiples

INTRODUCTION

Selon Economides (2004), les investisseurs sont dépassés par la quantité d'information à analyser et ne sont généralement pas en mesure d'en tirer pleinement avantage. Cette théorie n'est pas nouvelle. Chewning Jr et Harrell (1990) a démontré que les performances d'un individu s'améliorent initialement lorsque la quantité d'information augmente, mais diminuent rapidement lorsque celle-ci dépasse les capacités de traitement. Paredes (2003), Eppler et Mengis (2004) et Agnew et Szykman (2005) présentent sensiblement les mêmes constats.

Notre partenaire dans le cadre de ce projet, Croesus Finansoft, une entreprise de Laval spécialisée dans le développement d'un outil de gestion de portefeuille utilisé par les plus grandes firmes de courtage au pays, confirme le problème. Bien qu'ils ne conservent que le prix à la fermeture des marchés, les tables des historiques des titres dépassent les 500 millions d'entrées. Chez leur plus gros client, l'application gère 350 000 clients, 2 millions de comptes contenant plus de 10 millions de positions et plus de 750 millions de transactions, le tout accessible en temps réel. Le défi est donc de taille. L'entreprise aimerait tirer profit de ces informations pour offrir aux gestionnaires de portefeuille des outils d'aide à la décision qui leur permettrait d'effectuer des choix plus éclairés.

La visualisation d'information permet de tirer profit des capacités cognitives humaines pour analyser de grands ensembles de données complexes. Ce domaine de recherche en pleine effervescence trouve racine dans plusieurs autres domaines où de grands ensembles de données doivent être analysés, tels que la sécurité informatique, l'ingénierie logiciel, les sciences pures, la médecine, etc. La visualisation est également largement utilisée en finance, le plus souvent dans le cadre de l'analyse technique (Kirkpatrick et Dahlquist (2010)). Il existe toutefois très peu de cas documentés où la visualisation a été utilisée pour répondre à des problèmes plus complexes.

Croesus s'intéresse particulièrement à l'utilisation de la visualisation dans le contexte de l'analyse de performances. Contrairement à la croyance populaire, évaluer les performances d'un investissement ne consiste pas simplement à comparer les montants finaux et initiaux afin de cal-

culer le rendement obtenu. Le processus d'analyse se décompose en fait en trois problèmes distincts : la mesure, l'attribution et l'évaluation des performances (Lawton et Jankowski (2009)). Le premier consiste bien à mesurer le rendement net d'un investissement pour une période, mais la technique de calcul rend la compréhension des performances un peu plus complexe qu'une simple soustraction. En effet, les investisseurs utilisent habituellement une technique de calcul de rendement pondérée en fonction du temps (time-weighted rate of return). Avec cette technique, il est possible d'obtenir un rendement négatif même si ultimement un gain net a été enregistré. Cette situation est possible, par exemple, si une perte a été enregistrée pour la majeure partie du mois et que les gains ont été observés uniquement dans les derniers jours du mois. Le rendement sert donc à la fois d'indicateur de rendement et de niveau de risque. Le calcul des rendements doit ainsi prendre en compte les flux de trésoreries au cours du mois, les taux de changes, etc. À cause de tous ces éléments de complexité, même avec les calculs complets, les investisseurs ont souvent de la difficulté à bien comprendre les résultats obtenus.

Lors de l'étape d'attribution des performances, on compare les rendements des investissements à une référence, habituellement un indice de marché. L'objectif consiste à évaluer de quelle façon le rendement d'un investissement se compare au marché ou au secteur de marché dans lequel les titres du portefeuille évoluent. Il est évident qu'une perte de 5% sur un investissement ne réjouira personne. Toutefois, si le reste du marché a encaissé une perte de 25%, cet investissement a quand même mieux fait par rapport à la moyenne des autres investisseurs (20% de plus de capital restant à la fin de la période). Une multitude de facteurs doivent être pris en considération lors de l'attribution des performances : la force du marché, le contexte économique, les particularités du secteur d'industrie, la devise, l'inflation, etc. Des formules permettent de décomposer le rendement en facteurs afin d'identifier clairement la source des résultats obtenus (Bacon, 2008).

Effectuer l'analyse des performances sur un seul portefeuille ou pour un nombre limité de clients n'est pas trop complexe, moyennant une simple grille Excel et quelques formules mathématiques. Le défi devient toutefois de taille lorsque le nombre de clients dépasse le millier. Est-ce réellement réaliste de le faire au niveau de la firme, avec plus de 2 millions de comptes ?

Vu le degré de complexité et du nombre de paramètres impliqués, il existe très peu d'outils permettant d'analyser les performances d'une grande quantité de portefeuilles d'investissements.

0.1. Objectifs de la thèse

Les travaux de cette thèse consistent ultimement à proposer des techniques novatrices en visualisation permettant l'analyse des performances d'une quantité importante de portefeuilles. Les outils proposés doivent permettre d'identifier clairement les écarts entre les références utilisées et le portefeuille analysé, ainsi que les sources de ces écarts. Pour se faire, ils doivent faciliter l'identification des portefeuilles problématiques déviant de la norme (d'une référence) et de les décomposer, de façon à pouvoir identifier la ou les sources de problèmes ayant causé cette déviation.

Le principal défi du projet consiste à développer des techniques qui supportent la quantité de données à laquelle Croesus est confronté. L'analyse de performances est beaucoup plus intéressante lorsqu'elle permet d'identifier des problèmes cachés dans de grands ensembles de comptes, de façon à pouvoir prendre des actions correctives avant qu'il ne soit trop tard.

Dans le cadre de ce projet, un portefeuille correspond simplement à un ensemble d'investissements liés entre eux. Par exemple, l'ensemble des positions détenues par un investisseur peut être interprété comme un portefeuille. Dans le même ordre d'idées, l'ensemble des investissements d'une firme peut être considéré comme le portefeuille de la firme. L'analyse des performances peut être effectuée à tous les niveaux.

Les portefeuilles peuvent également être perçus comme des structures hiérarchiques. Par exemple, à la cime de l'arbre se trouve la firme. Cette dernière peut généralement être décomposée en plusieurs régions (ex : est/ouest du pays). Dans chaque région se trouvent des succursales ou des bureaux. Chaque succursale emploie à son tour un ou plusieurs conseillers qui gèrent les portefeuilles de plusieurs clients. Chaque client peut également diviser ses investissements en plusieurs portfolios. Il serait possible d'imaginer encore davantage de niveaux à cette structure.

Chaque niveau peut être perçu comme un portefeuille contenant un ou plusieurs sous-portefeuilles. Cette structure hiérarchique est particulièrement intéressante dans la mesure où elle permet de décomposer des investissements. L'outil devrait mettre en valeur cette réalité et permettre d'explorer les différentes couches de façon à mieux comprendre les résultats obtenus et possiblement de permettre de descendre jusqu'à la source de l'anomalie (ex : le titre ayant eu un impact sur le portefeuille de la succursale).

Il est primordial de prendre en considération le public cible. Les systèmes proposés doivent être suffisamment simples pour être utilisés dans un environnement de production par des spécialistes en finance, une clientèle qui est réputée pour être réticente aux changements trop brusques. Par conséquent, les techniques devront être validées auprès d'experts afin de s'assurer qu'ils répondent réellement aux besoins et qu'ils sont accessibles par la clientèle ciblée. Évidemment, une formation minimale pourrait être nécessaire afin de bien maîtriser les outils proposés. Croesus offre déjà ce genre de formation lors de la mise en production des nouvelles versions de ses produits.

Le projet ne consiste pas à développer une application complète permettant d'effectuer de l'attribution de performances. Il s'agit plutôt d'explorer différentes techniques permettant de faciliter l'exploration et l'analyse des performances financières. Les contributions pourraient éventuellement être intégrées à une application développée par Croesus pour effectuer l'analyse détaillée des performances. Les contributions proposées prennent d'ailleurs la forme, pour la plupart, de graphiques, de contrôles ou de techniques d'interaction novateurs permettant de faciliter l'analyse et l'exploration des données financières.

Puisque les modèles d'attribution diffèrent en fonction des classes d'actifs évaluées, l'attribution sera optimisée pour les portefeuilles composés d'actions et de fonds, qui composent la plus grande partie des portefeuilles gérés par Croesus. Bien qu'il serait intéressant de supporter d'autres types de portefeuilles, par exemple à revenus fixes (ex : obligations d'épargne), l'entreprise ne possède pas les données nécessaires pour effectuer ce type d'analyse.

0.2. Méthodologie

Le projet de recherche a été développé dans un cadre de partenariat avec Croesus Finansoft. Plus de 50% du temps de recherche a été passé en entreprise, de façon à pouvoir réaliser, en collaboration avec les experts de l'entreprise, des solutions innovantes qui répondent réellement aux besoins de l'industrie.

Des analystes d'affaire de Croesus, possédant des connaissances avancées en finance et ayant pour la plupart été antérieurement conseillers eux-mêmes, ont été impliqués dans le projet dès le début afin d'établir clairement les besoins. Ils ont également été sollicités à toutes les étapes charnières afin de valider les prototypes, effectuer certaines expériences, répondre à des questions techniques, etc.

Une revue de la littérature exhaustive a été effectuée pour identifier les idées les plus prometteuses. Les résultats ont d'ailleurs été publiés sous la forme d'une affiche lors de la première conférence internationale en visualisation (Dumas *et al.*, 2014) en octobre 2014, afin de maximiser la portée du travail et de s'assurer que les résultats étaient accessibles par tous les professionnels du milieu. Les résultats sont discutés dans le chapitre 1 et présentés en détail en annexe A.

À partir des informations recueillies auprès des analystes, différents prototypes ont été réalisés et ont été discutés avec l'équipe de direction de l'entreprise. Un développement itératif s'en est suivi, jusqu'à ce que les résultats répondent aux objectifs établis.

Afin de valider les résultats obtenus, les prototypes ont été soumis à des processus d'évaluations complets. Le type d'évaluation dépend évidemment de la technique sous évaluation. Dans deux des projets, il s'agissait d'expériences contrôlées en laboratoire visant à évaluer comment la technique se compare à l'état de l'art. Le dernier projet a plutôt été évalué avec un panel d'experts du domaine qui ont passé près de 3 heures chacun à explorer en détail la solution proposée dans un contexte d'analyse réel avec des données provenant d'une grande firme canadienne.

0.3. Organisation du document

Le chapitre 1 présente une revue détaillée de la littérature commune à tous les projets. Une revue plus spécifique pour chacun des projets est présentée directement au début des chapitres 2, 3 et 4.

Le chapitre 2 présente VectorLens, un outil permettant d'explorer de grands ensembles de données de performances présentés sur un graphique linéaire simple. Cet outil a été créé dans le contexte où un analyste désirerait comparer les performances de plusieurs centaines (voir milliers) de portefeuilles (ou de titres) entre eux. Bien que les graphiques soient très chargés, il est possible de voir ressortir des tendances globales et des cas anormaux. VectorLens simplifie l'analyse en permettant de sélectionner uniquement les titres présentant les patrons à investiguer à l'aide d'une technique novatrice de sélection angulaire en un seul glissement. L'outil permet également d'obtenir des informations complémentaires sur la sélection résultante et d'appliquer des filtres supplémentaires pour raffiner la sélection. Une analyse complète de l'outil est présentée dans le chapitre, en plus des détails de l'évaluation de la technique par rapport à l'état de l'art. Cet article a été publié dans le journal Transactions on Visualization and Computer Graphics en mars 2015 (Dumas *et al.*, 2015).

Le chapitre 3 présente les couches compressées (compressed layers), une technique novatrice permettant d'optimiser la disposition des couches de données sur un graphique linéaire simple. Dans un contexte similaire à VectorLens, on tente d'analyser plusieurs regroupements de portefeuilles, par exemple les performances annuelles des portefeuilles de deux ou trois gestionnaires de comptes. Il serait possible de séparer les couches (les gestionnaires) à l'aide de couleurs différentes. Toutefois, vu la quantité de courbes impliquées, il devient très difficile d'identifier clairement les couches à cause de l'occlusion résultante. La compression consiste à optimiser l'espace sous-exploité entre deux points (deux mois, par exemple) sur l'axe des X en compressant à des niveaux différents chaque couche. Les résultats d'une étude en laboratoire sont également présentés pour appuyer cette approche.

Le chapitre 4 présente deux techniques novatrices basées sur le graphique ternaire (ternary plot) : la roue ternaire et la barre de navigation ternaire. Les deux techniques ont été développées afin de représenter sur un même graphique le rendement différentiel avec une référence et les effets expliquant cet écart pour un grand nombre de portefeuilles. Les deux techniques ont également été introduites dans un système complet d'analyse de performances. Ce système a ensuite été évalué par quatre experts du domaine dans une étude de cas utilisant des données réelles de production.

CHAPITRE 1

REVUE DE LA LITTÉRATURE

Cette section présente les concepts généraux communs à tous les articles. Une revue de la littérature plus ciblée est présentée dans chacun des chapitres pour les éléments particuliers propres à chaque article.

1.1 La visualisation de l'information

Selon Thomas et Cook (2005), la visualisation d'information peut être décrite comme la science du raisonnement analytique, facilité par des interfaces visuelles interactives. Spence (2006) décrit la visualisation comme l'action de former une image ou un modèle mental d'un concept. McCormick *et al.* (1987) est le premier à définir la visualisation comme une science informatique. Il décrit la visualisation comme suit :

> Visualization is a method of computing. It transforms the symbolic into the geometric, enabling researchers to observe their simulations and computations. Visualization offers a method for seeing the unseen. It enriches the process of scientific discovery and fosters profound and unexpected insights. (McCormick *et al.*, 1987, p. 3)

Pour Munzner (2015), un système de visualisation permet d'offrir une représentation visuelle d'un ensemble de données de façon à permettre aux utilisateurs d'effectuer certaines tâches plus efficacement. L'humain joue un rôle critique dans le processus décisionnel. Contrairement aux systèmes automatisés, la visualisation intègre l'humain dans le processus en lui présentant les informations de manière stratégique et en le guidant dans le processus d'exploration des données. Cette approche a l'avantage de ne pas nécessiter la définition de règles, et permet de découvrir des patrons ou des anomalies inattendues pour lesquels il aurait été difficile d'établir des règles au préalable. L'expert du domaine a également l'avantage, grâce à ses connaissances et à son expérience, de mieux comprendre les données dans leur contexte, ce qui permet, par

exemple, de limiter les cas de faux positifs, particulièrement dommageables avec les approches automatisées.

La visualisation n'est pas nécessairement l'outil de prédilection dans tous les cas. S'il est possible de définir une tâche de façon précise, de manière à ce que des statistiques ou des algorithmes d'apprentissage machine puissent répondre efficacement et correctement aux questions, il n'est pas nécessaire d'impliquer l'humain dans le processus décisionnel. Toutefois, plusieurs problèmes sont difficiles à définir avec précision. Selon Munzner, dans ces circonstances, la meilleure approche consiste à exploiter les capacités avancées du système visuel humain à détecter des patrons dans les données. La visualisation permettra alors d'augmenter les capacités de l'humain plutôt que d'essayer de le remplacer complètement.

Plus précisément, selon Lemieux *et al.* (2014), la visualisation est particulièrement intéressante pour les tâches analytiques qui demandent :

– De traiter de grandes quantités de données complexes.
– De répondre à une quantité importante de questions souvent ambiguës.
– De conserver l'humain au centre du processus d'analyse.
– De mélanger l'analyse automatisée avec des visualisations interactives.
– De fournir une réponse rapide avec des détails sur demande.
– De présenter les résultats de l'analyse.

La visualisation peut être décrite comme une science tridisciplinaire qui repose à la fois sur les technologies de l'information, la psychologie et la biologie (plus particulièrement le système visuel et la perception). Une bonne compréhension de ces trois domaines est critique pour la réalisation de systèmes de visualisation fonctionnels et efficaces.

Selon Ware (2013), les yeux et le cortex visuel forment un processeur massivement parallèle offrant une bande passante très importante à l'humain pour la perception. Selon l'auteur, en déchiffrant le fonctionnement de la perception humaine, il est possible d'utiliser ces connaissances pour former des représentations graphiques permettant aux éléments importants des

données de ressortir par eux-mêmes. Pour Lemieux *et al.* (2014), ce sont les capacités spatiales évoluées de l'humain qui lui permettent d'interpréter efficacement une représentation visuelle.

> Visualization had advantages over other modes of communication because humans have evolved visual and spatial skills that include the ability to detect edges and discontinuities, things that stand out, variations in color and shape, and motion ; to recognize patterns ; and to retrieve information using visual cues. (Lemieux *et al.*, 2014, p. 286)

La visualisation présente une multitude d'avantages. La visualisation permet d'étendre les capacités de raisonnement de l'humain de six façons (Thomas et Cook (2005); Card *et al.* (1999); Lemieux *et al.* (2014)) :

– En augmentant les ressources cognitives, de façon à ce que les ressources visuelles puissent augmenter la mémoire fonctionnelle humaine.

– En réduisant la recherche, par exemple en présentant de grandes quantités de données dans un petit espace.

– En augmentant les capacités à reconnaître des patrons, par exemple en structurant les données dans le temps ou dans l'espace.

– En facilitant l'inférence des relations entre les éléments (qui seraient difficiles à détecter autrement).

– En analysant perceptuellement une grande quantité d'événements potentiels.

– En fournissant un médium manipulable qui, contrairement à des diagrammes statiques, permettent d'explorer tout l'espace des paramètres possibles.

Selon Ware, les principaux avantages de la visualisation peuvent être résumés de la façon suivante :

– La visualisation permet d'interpréter des quantités très importantes de données. Il est possible d'afficher les informations importantes de plus d'un million d'éléments sur une seule visualisation compréhensible.

– La visualisation permet la perception de propriétés émergentes non anticipées. Elle permet de faciliter la perception de patrons dans les données qu'il serait très difficile d'obtenir autrement.

– La visualisation met en évidence les problèmes avec les données, par exemple avec la collecte des données.

– La visualisation permet de comprendre les caractéristiques des données à petite et à grande échelle. Elle permet de mieux comprendre les relations qui unissent les données.

– La visualisation facilite la formulation d'hypothèses. En observant les données, l'humain détecte des éléments intéressants ou inattendus qui demanderont d'autres analyses plus approfondies.

Ces caractéristiques font en sorte que l'utilisation de la visualisation s'adapte bien à une multitude de domaines. On retrouve la visualisation dans des domaines aussi variés que la médecine, l'art, la finance, la sécurité informatique, etc.

1.1.1   Les séries temporelles

Le temps peut à la fois être représenté comme un point dans l'histoire, comme un intervalle entre deux points, comme une série de points à intervalles réguliers, ou même comme une série continue de moments. Il peut représenter une séquence d'événements subséquents ou même des cycles. Le concept même du calendrier, avec des mois qui ne contiennent pas le même nombre de jours, par exemple, justifie à lui seul une considération particulière pour cette variable pourtant critique pour la plupart des systèmes de visualisation.

Aigner *et al.* (2011) propose une liste de critères à considérer lorsqu'il est question de représenter le temps dans une visualisation :

– **Échelle** : Est-ce que le domaine des données est continu (il existe une valeur pour tous les points possibles entre deux dates), discret (il est possible de comparer quantitativement des moments distincts discontinus dans le temps - ex : UNIX Timestamp) ou ordinal (il est uniquement possible de trier des événements dans le temps sans les comparer quantitativement)

– **Étendue** : Est-ce qu'il s'agit d'un moment discret dans le temps (point-based) ou d'une période continue (interval-based) ?

– **Arrangement** : Est-ce qu'il s'agit d'une séquence linéaire d'événements (du passé vers le futur) ou d'un cycle (ex : saisons, journées, etc.) ?

– **Point de vue** : Est-ce qu'il s'agit d'une séquence ordonnée d'événements, d'une arborescence d'événements (plusieurs événements qui surviennent simultanément) ou de plusieurs perspectives d'un même événement (plusieurs vues possiblement contradictoires d'un même événement - ex : rapports de témoins oculaires d'un crime).

– **Granularité** : Est-ce que le temps est représenté par un calendrier (multiple granularities - peut être représenté en mois, en jours, en heures, etc.), par une seule unité commune (single granularity - ex : millisecondes) ou par des points abstraits (pas associés à une unité en particulier).

– **Primitives temporelles** : Est-ce qu'il s'agit d'un seul point dans le temps (selon l'étendue, peut avoir une durée), un intervalle (constitué d'un point de début et d'une fin) ou une période (similaire à l'intervalle, mais n'est pas accroché à un moment dans le temps - ex : 4 jours). Les primitives sont reliées par des relations entre elles (ex : avant/après, égal, se chevauchent, etc.

– **Incertitude** : Est-ce que les informations sont certaines (determinate) ou incertaines (indeterminate - ex : peut prendre 2 ou 3 semaines).

De façon générale, les données financières sont présentées sous forme de séquences à intervalles réguliers (à chaque minute, à chaque jour, etc.). Certaines sources peuvent également fournir des données en temps réel. Dans tous les cas, il s'agit d'un ensemble de valeurs (ex : prix des titres, métriques, etc.) associées à un moment dans le temps. Cette séquence de données porte le nom de série temporelle (time series).

Il existe une multitude de techniques permettant de visualiser des séries temporelles. Tominski et Aigner (2013) présente un survol d'une centaine de techniques de visualisation de séries temporelles, regroupées entre autres par arrangement et par primitives temporelles. Les résultats du survol sont publiés sur un site web accessible à l'adresse http ://www.timeviz.net/. La figure 1.1 présente une capture d'écran du site. Il s'agit d'ailleurs de l'une des inspirations du projet FinanceVis.net présenté en annexe A.
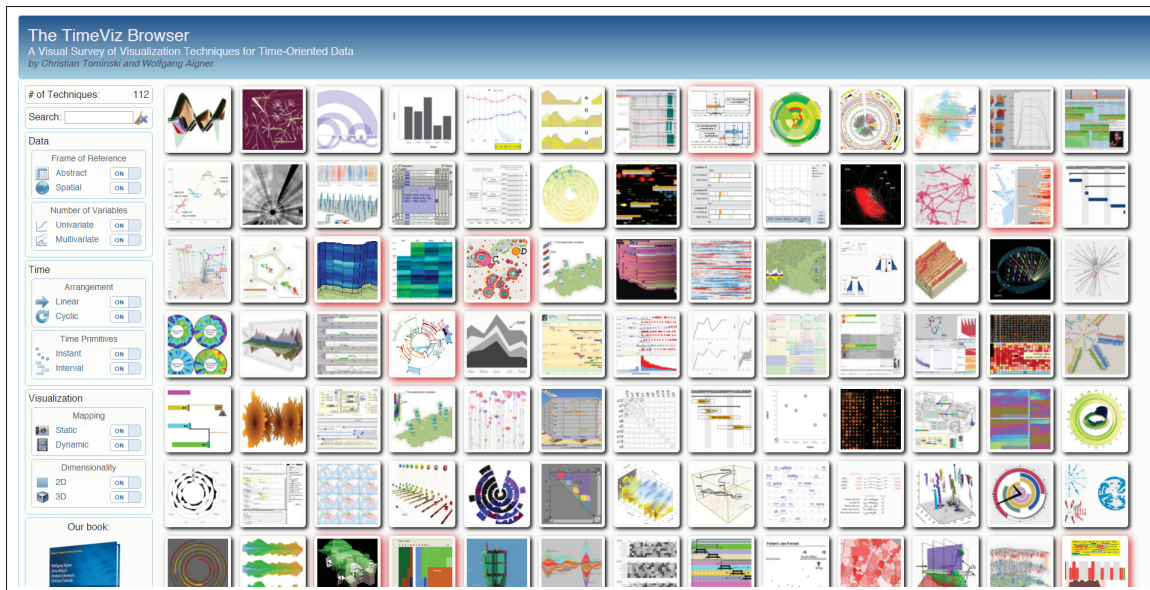
Figure 1.1   Interface du navigateur de visualisations de séries temporelles réalisé par
Tominski et Aigner (2013).

1.1.2   Les interactions

Selon Spence (2006), les trois ingrédients d'une bonne visualisation sont : la représentation visuelle, la présentation et les interactions. Une vue statique peut certes présenter l'essentiel du message selon des paramètres prédéfinis. Toutefois, en permettant à l'utilisateur de changer les paramètres de la vue, de sélectionner directement des éléments ou de naviguer à travers l'ensemble des données, l'utilisateur peut obtenir une vue beaucoup mieux adaptée à ses besoins et obtenir des compléments d'information qui ne seraient pas possibles d'obtenir à l'aide d'une simple vue statique.

Il existe une multitude d'interactions possibles selon le médium utilisé. Yi *et al.* (2007) proposent une classification générique des principales techniques d'interaction avec des visualisations.

– **Abstract/Elaborate** : Afficher plus de détails (ex : "drill-down", infobulles, zoom)
– **Connect** : Afficher des items en relations (ex : vues coordonnées, surbrillance)
– **Encode** : Afficher une représentation différente (ex : couleur, taille, forme, orientation)

– **Explore** : Afficher quelque chose d'autre (ex : examiner un autre sous-ensemble de données)

– **Filter** : Afficher quelque chose de façon conditionnelle

– **Reconfigure** : Afficher un arrangement différent (ex : tri, réogranisation des colonnes ou des axes, rotation de la vue)

– **Select** : Identifier un élément de la visualisation comme "intéressant"

– **Undo/redo** : Permettre de retourner à un état antérieur

– **Change configuration** : Permettre d'adapter l'interface à ses besoins

Cette classification a été utilisée dans le cadre du projet FinanceVis.net (voir annexe A) afin de catégoriser les types d'interactions supportées par les différentes visualisations.

Il existe plusieurs stratégies permettant de présenter de grands ensembles de données à l'utilisateur sans le surcharger. La technique la plus classique porte le nom de "Information Seeking Mantra" (Shneiderman, 1996). Elle dicte trois règles que devraient respecter le concepteur d'une visualisation : "overview first, zoom and filter, then details-on-demand.". Selon le mantra, la visualisation devrait afficher un aperçu de l'ensemble des données à l'utilisateur dès le départ et offrir des techniques d'interaction permettant d'approfondir uniquement ce qui est intéressant, sur demande. Keim *et al.* (2006) a proposé une variante du mantra : "Analyze First, Show the Important, Zoom, Filter and Analyse Further, Details on Demand.". Cette variante met l'emphase sur l'importance de l'analyse des données dans le processus, mais reflète les mêmes règles de conception. L'approche opposée a également été proposée par van Ham et Perer (2009) : "Search, Show Context, Expand on Demand". van den Elzen et van Wijk (2014) ont également proposé à la blague "the NAMREDIENHS mantra" (reverse Shneiderman mantra) : "Details, Select & Aggregate, Overview". Tous les projets réalisés dans le cadre de cette thèse utilisent l'approche de Shneiderman, en présentant l'ensemble des données à haut niveau et permettent, via le filtrage, le "drill-down", des infobulles ou autres, d'obtenir plus de détails.

Il est fréquemment nécessaire de présenter des informations à plusieurs niveaux. Par exemple, dans le cadre de l'analyse des performances, on cherche à représenter une hiérarchie de porte-

feuilles. Chaque niveau correspond à l'agrégation des niveaux inférieurs (ex : portefeuille ->
client -> conseiller -> succursale -> province -> firme). Plusieurs techniques ont été proposées
pour traiter des données à plusieurs niveaux (généralement deux à la fois). Ces stratégies sont
regroupées sous les catégories "Overview+detail" et "focus+context" (Aigner *et al.*, 2011).
"Overview+detail" présente le survol et les détails dans deux vues séparées, alors que "fo-
cus+context" intègre les deux concepts dans une même vue. Cockburn *et al.* (2009) présente
un survol détaillé de différentes techniques pour les deux catégories. La plupart des techniques
impliquant des lentilles (voir chapitre 2) sont considérées comme focus+context.

Dans le cas où l'interface présente plusieurs vues, Wang Baldonado *et al.* (2000) recommende
d'utiliser des vues coordonnées, où la sélection d'un élément sur une vue se reflète sur les
autres de façon synchronisée. Cette technique permet à l'utilisateur de bien comprendre le lien
entre les différentes vues et de repérer plus facilement les éléments communs. Le dernier projet
de cette thèse (chapitre 4) utilise ce concept entre les différentes vues coordonnées.

1.1.3 Évaluer une visualisation

L'évaluation d'une visualisation permet de valider les hypothèses de conception établies initia-
lement et de fournir des preuves qu'elles sont fondées (Hoeber, 2009). Elles permettent égale-
ment de comparer différents systèmes ou techniques de visualisation entre eux. Selon Plaisant
(2004), les évaluations identifient clairement le potentiel et les limites des outils. Pour Kosara
et al. (2003), elles permettent également de prouver qu'elles sont utiles dans un sens pratique
pour des tâches précises, et surtout pourquoi elles sont efficaces. Il est également possible de
valider des concepts théoriques abstraits de façon à identifier les conditions où ces théories sont
valides.

Il existe une multitude de techniques permettant d'évaluer un système de visualisation. Les
techniques les plus courantes sont :

– **Expériences en laboratoire (controlled experiment)** : Études réalisées par des participants
   volontaires dans un environnement contrôlé. Les participants n'ont pas besoin d'être des

experts du domaine. Permet d'évaluer efficacement aussi bien des aspects perceptuels de bas niveau que des systèmes complets de visualisation (Hoeber, 2009). Génère des métriques quantitatives, telles que le taux d'erreurs, le niveau de précision, le temps nécessaire, etc. Il s'agit de la technique la plus fréquemment utilisée pour valider des systèmes de visualisation. Les chapitres 2, 3 et 4 présentent des projets ayant recours à des études en laboratoire pour démontrer l'efficacité des techniques ou pour mieux comprendre les caractéristiques des solutions comparées.

– **Études de cas** : Études réalisées généralement avec des experts du domaine. Le système de visualisation est évalué idéalement dans un contexte réel avec de vraies données. Cette technique a l'avantage d'évaluer le travail de l'expert dans son environnement de travail avec de véritables tâches d'analyse (Plaisant, 2004). Les résultats sont toutefois plus difficiles à reproduire. Ces études produisent également des résultats plus qualitatifs. Des études de cas sans expert (données analysées par les concepteurs de la visualisation pour démontrer l'utilisation) sont généralement moins bien reçues par la communauté.

– **Évaluation heuristique ou évaluation d'experts** : Des experts utilisent une liste d'heuristiques prédéfinies pour évaluer l'interface proposée. Génère une liste de commentaires et d'observations. Permet d'identifier les problèmes qui pourraient surgir lors de l'utilisation de l'interface. Peut également prendre la forme d'un "cognitive walk-through", où les experts passent à travers une liste de tâches et tentent de trouver tous les problèmes qui pourraient survenir à chaque étape du processus (Tory et Möller, 2004).

– **Entrevues avec des experts / focus groups** : Rencontres individuelles ou en groupes avec des experts du domaine. Les experts doivent répondre à des questions prédéfinies visant à générer des idées ou à valider des hypothèses.

– **Études longitudinales** : Évalue l'outil dans un contexte de production durant plusieurs mois / années. Bien que ces études produisent les résultats les plus rigoureux, elles demandent des efforts considérables et sont par le fait même moins fréquentes.

De façon générale, les évaluations formelles sont effectuées à la fin du projet, lorsqu'un prototype fonctionnel de la visualisation ou du système a été développé. Toutefois, selon Munzner

(2008), cette validation devrait être effectuée tout au long du processus de conception, de façon à corriger la situation dès que possible. Munzner propose d'ailleurs un cadre de conception et d'évaluation d'une visualisation avec des tâches spécifiques à réaliser à différentes étapes charnières du processus. La figure 1.2 présente une formalisation du modèle proposé par Munzner.
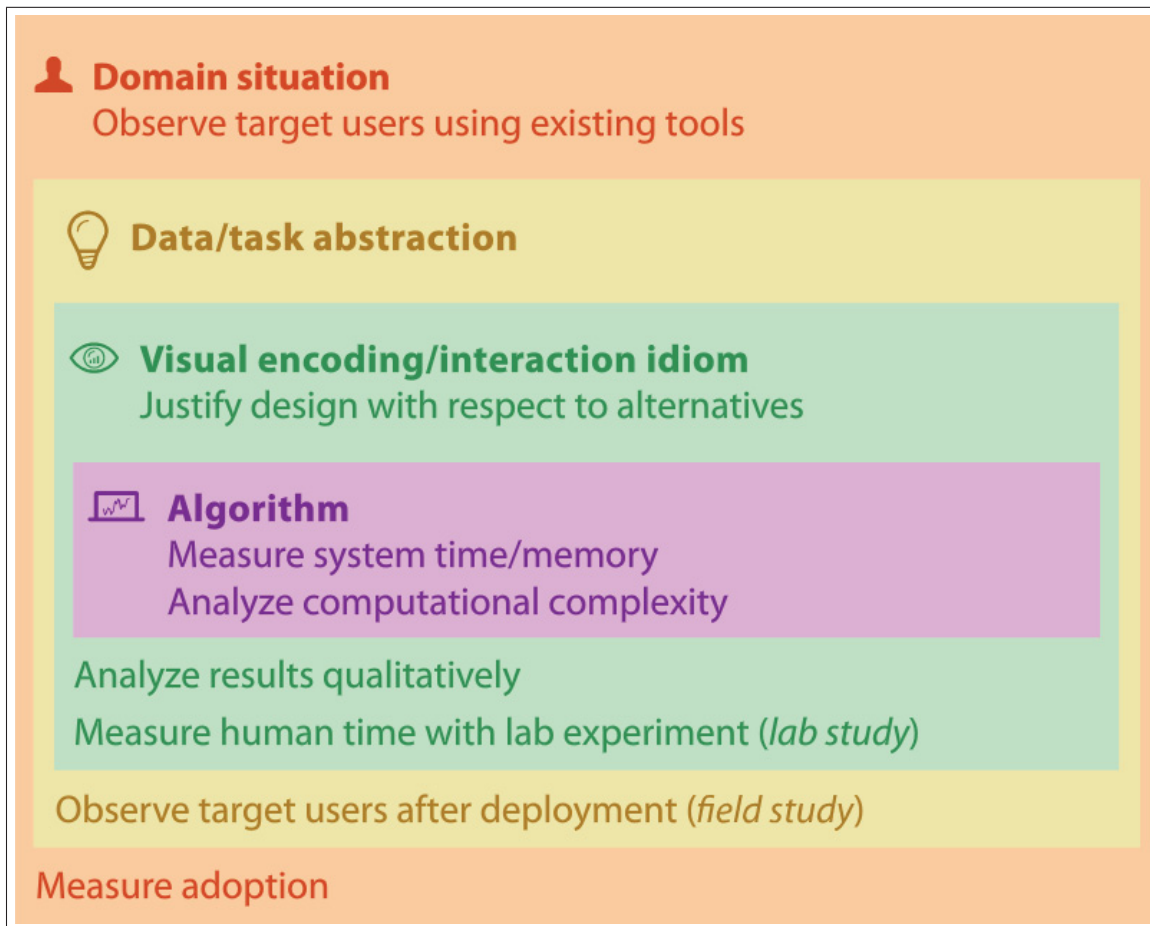
Figure 1.2   Le modèle imbriqué de conception et d'évaluation de Munzner.
Tirée de Munzner (2015).

La première étape du processus de conception consiste à comprendre la situation du domaine, le vocabulaire utilisé, les flux de travaux (workflows), comment les experts utilisent les données à leur disposition pour effectuer leur travail, etc. Le principal risque à cette étape, selon Munzner, est d'identifier le mauvais problème ou de mal le caractériser. Une technique pour valider les hypothèses initiales consiste à conduire des entrevues avec des experts, à observer leurs tâches

et à identifier clairement les exigences du projet en partenariat avec les experts. Cette étape sera également évaluée à la fin du processus en mesurant le taux d'adoption de l'outil auprès de la clientèle ciblée.

L'étape suivante consiste à effectuer une abstraction plus générique des données et des questions propres au domaine. On établit alors une liste de tâches génériques de visualisation ciblées par le système à concevoir. Par exemple, le système peut permettre de naviguer dans les données, de comparer des ensembles de données, etc. Cette étape sera évaluée après le déploiement du système par une étude sur le terrain par les experts dans le cadre de leur travail. Une étude longitudinale serait une technique intéressante dans ce contexte.

La troisième étape vise à identifier les techniques permettant de représenter visuellement les informations du domaine et permettant de répondre aux tâches demandées. On détermine également les techniques d'interaction permettant de manipuler les informations et les représentations. Il est possible d'utiliser des évaluations heuristiques et des évaluations par des experts pour justifier les techniques sélectionnées. Une fois le développement terminé, une étude en laboratoire permettra de confirmer empiriquement les hypothèses.

Le niveau le plus imbriqué concerne le choix et la conception des algorithmes permettant à l'ordinateur de générer les résultats attendus. Le principal défi à ce niveau concerne les performances et l'utilisation mémoire. Il est possible d'évaluer la complexité de calcul avant d'implémenter l'algorithme, et de mesurer les performances par la suite afin de s'assurer qu'il respecte les objectifs établis.

1.2 La visualisation de données financières

Bien que les marchés d'échanges existent depuis 2000 ans avant J-C. (Kirkpatrick et Dahlquist (2010)), il faut attendre au 18e siècle avant de voir apparaître les premières visualisations en finance. Selon (Nison (2003)), les Japonais ont développé à cette époque des méthodes d'analyse technique basée sur les chandeliers japonais (figure 1.3).
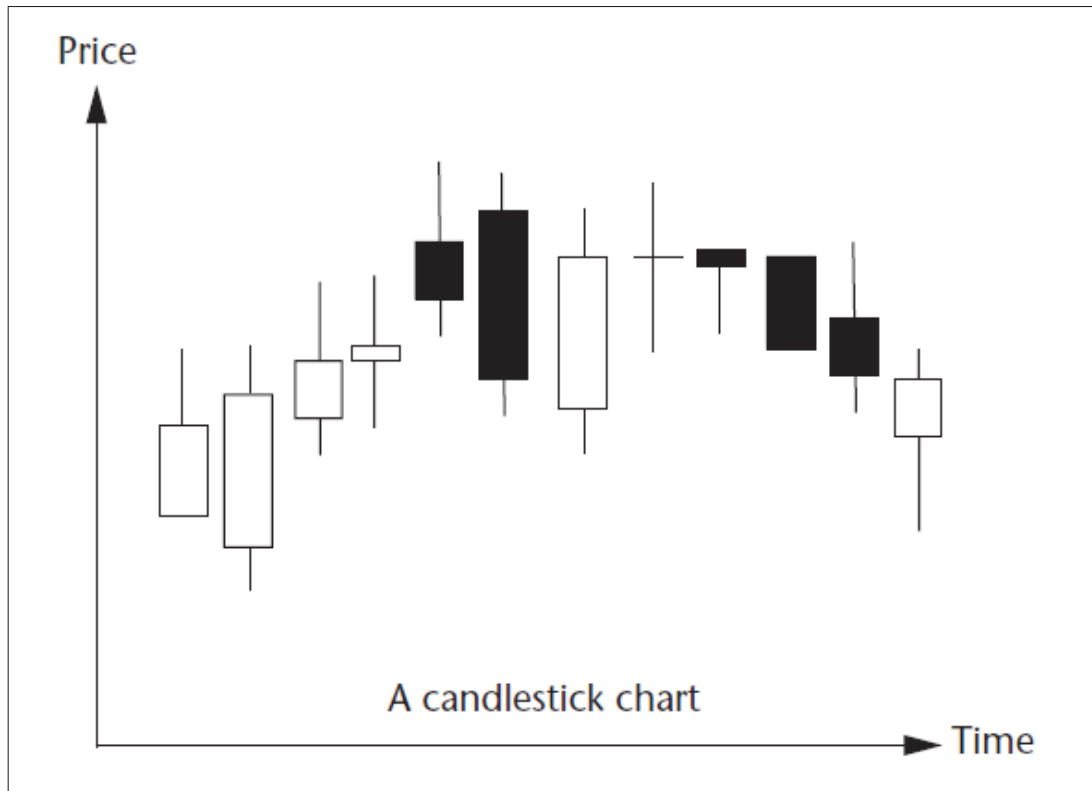
Figure 1.3   Exemple de diagramme utilisant les chandeliers japonais.
Tirée de Nesbitt et Barrass (2004).

On entend par analyse technique "l'étude des prix dans un marché de libre échange dans l'objectif d'effectuer des échanges profitables" (Kirkpatrick et Dahlquist (2010)). Un analyste technique base uniquement ses décisions sur l'historique des prix d'un titre. Cette logique repose sur l'hypothèse que le marché est toujours efficient (les prix et rendements des actifs reflètent toutes les informations disponibles) et que les différents facteurs qui influencent l'offre et la demande sont toujours pris en considération dans le prix de l'action. Outre les chandeliers japonais, les analystes financiers utilisent habituellement des graphiques présentant l'évolution des prix des titres combinée à différentes métriques (ex : moyennes mobiles, oscillateurs, bandes Bollinger, etc.). La figure 1.4 illustre un exemple de graphique utilisé en analyse technique.

Outre pour représenter l'évolution des prix des titres, la visualisation est généralement très peu utilisée en finance. À peine quelques produits commerciaux offrent des visualisations intéressantes ne faisant pas appel à l'analyse technique. Par exemple, le magazine SmartMoney a

Figure 1.4   Exemple de diagramme utilisé en analyse technique. On y retrouve, dans ce cas-ci, les courbes des moyennes mobiles de 20 (rouge) et 50 jours (vert).
Tirée de Google (2013).

développé Map of the Market, une solution diffusée sur leur site web et présentant les titres les plus actifs sur le marché, regroupés par secteur d'activité, à l'aide d'un treemap (Wattenberg (1999)). La figure 1.5 présente la carte du marché du 19 mars 2013. La couleur représente un gain (vert) ou un perte (rouge), et l'opacité le pourcentage de variation. La taille de la case représente la capitalisation boursière.

L'entreprise Gravity Investments se démarque également du lot avec son produit GSphere, permettant de projeter sur une sphère 3D le degré de diversification du portefeuille basé sur plusieurs métriques complexes. La figure 1.6 illustre le résultat d'une bonne diversification avec l'outil. Market Topology (Impactopia (2013)) offre également une projection 3D des variations des titres sur le marché. Outre ces solutions, la plupart des solutions commerciales se limitent généralement à des visualisations simples, comme des courbes de prix, des histogrammes, des diagrammes en pointe de tarte, etc.

Du côté académique, plusieurs contributions dans le domaine de la visualisation ont été publiées au cours des dernières années dans les conférences et les journaux de visualisation ou de finance. Nous avons réalisé un survol exhaustif de la littérature et publié nos résultats sous
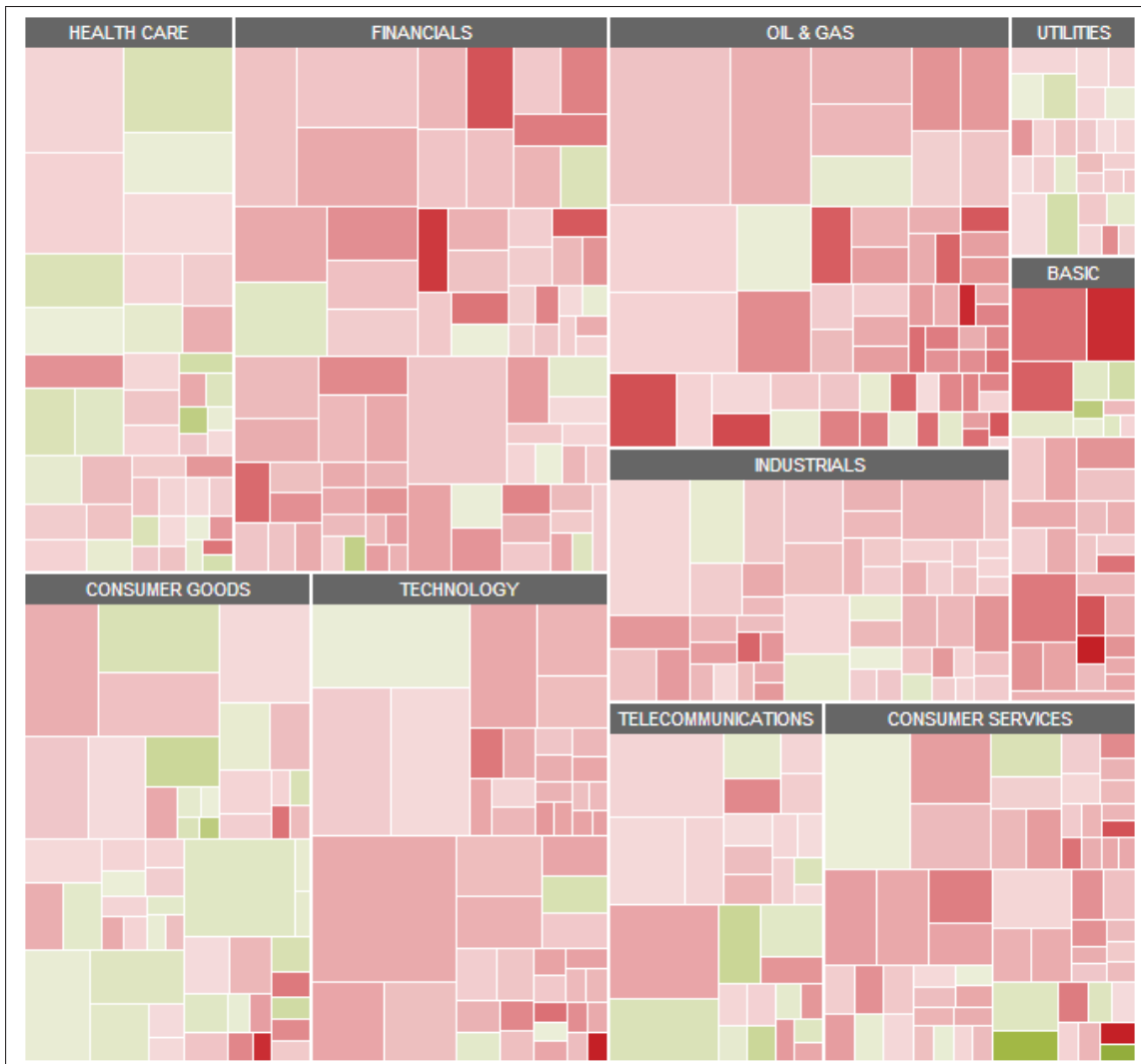
Figure 1.5   Carte du marché du 19 mars 2013. On remarque par la dominance de la
couleur rouge que le marché était fortement en baisse cette journée-là.
Tirée de SmartMoney (2013).

la forme d'un site web lors de la conférence IEEE Vis 2014 (Dumas *et al.*, 2014). L'annexe A
présente l'article court accompagnant l'affiche publiée lors de la conférence.

La figure 1.7 présente un résumé de la contribution. Plutôt que de regrouper les visualisations
par tâche financière, nous avons présenté une structure hiérarchique à trois niveaux : la topo-
logie, la dimension et la mise en correspondance (mapping) des données. Au premier niveau,
les visualisations sont regroupées en fonction des relations entre les données, soit s'il s'agit
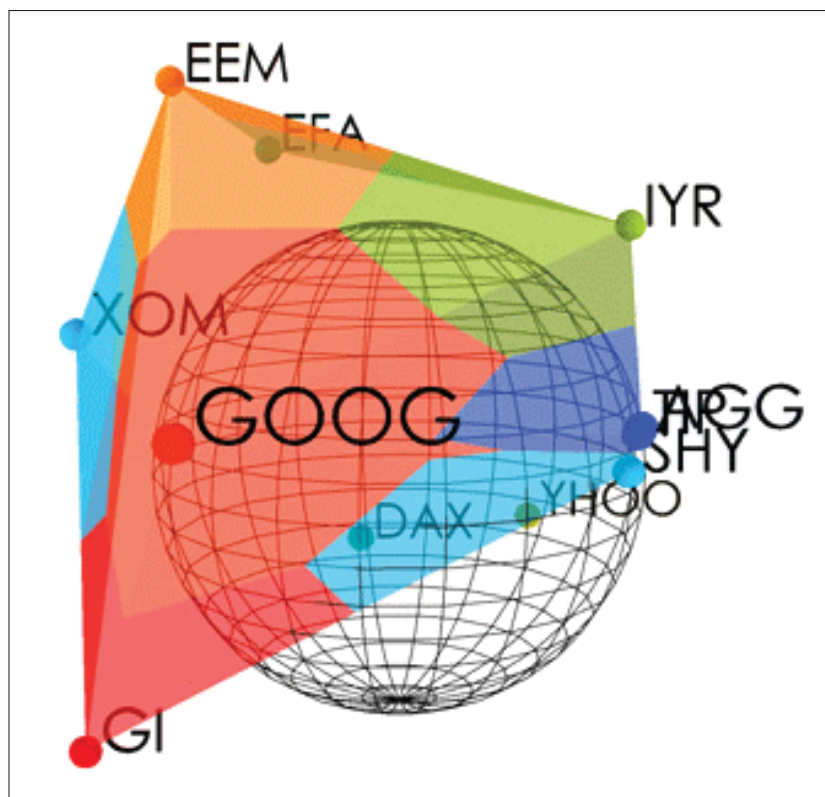
Figure 1.6   Portefeuille diversifié composé de 11 titres
visualisé avec G-Sphere.
Tirée de Investments (2013).

d'une arborescence, d'un réseau ou autre (sans topologie intrinsèque). On divise ensuite les trois groupes selon s'il s'agit de représentations 2D ou 3D, et si la représentation est statique ou dynamique (si les données changent durant l'analyse - ex : systèmes temps réels). Nous croyons que cette taxonomie permet de bien diviser l'espace et de faciliter la réutilisation des concepts en fonction des propriétés intrinsèques des données et des exigences de conception. Une visualisation qui a été développée pour visualiser des fonds mutuels pourrait très bien être utilisée dans un autre contexte, avec un autre ensemble de données complètement différent. La division proposée permet de trouver plus facilement des alternatives pour des données ayant des caractéristiques similaires.

Un site web a également été réalisé afin de diffuser les résultats (http://www.financevis.net). Le site présente un résumé complet de l'ensemble des caractéristiques de chacune des vues

et permet d'effectuer des recherches avancées selon les critères souhaités. Par exemple, il est possible de rechercher des techniques de visualisation par type et par nombre de variables dépendances et indépendantes, par type de visualisation, par dimension, par alignement (radial, perpendiculaire, libre), par algorithme d'intelligence machine et par techniques d'interaction (selon la taxonomie de Yi *et al.* (2007)). Les visualisations ont également été regroupées par tâche selon une taxonomie basée sur celle proposée par Lemieux *et al.* (2011).
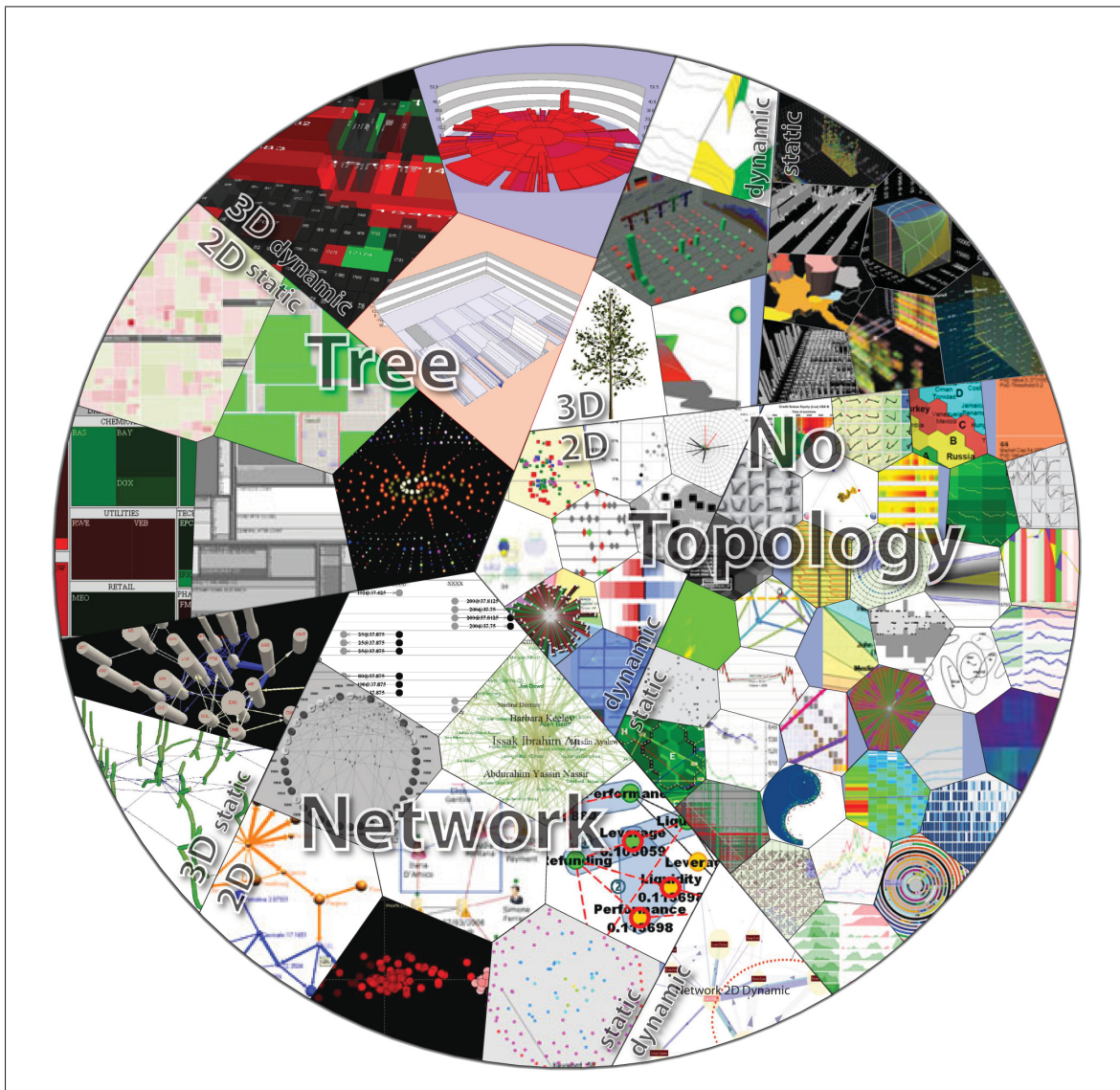
Figure 1.7   Extrait de l'affiche présentée lors de la conférence IEEE Vis 2014.

1.3 L'analyse de performances financières

Contrairement à la croyance populaire, analyser des performances financières ne consiste pas simplement à regarder le rendement d'un investissement. Il s'agit en fait uniquement de la première étape d'un processus relativement complexe et structuré, composé de trois étapes : la mesure, l'attribution et l'évaluation des performances (Lawton et Jankowski, 2009).

La mesure de performances est sans contredit l'étape la plus simple et la plus intuitive du processus. Il est toutefois important de considérer les flux monétaires durant la période évaluée lors du calcul du rendement. Il existe deux techniques permettant d'effectuer ce calcul : le rendement pondéré du coût de l'argent (money-weighted return - MWR) et le rendement pondéré par un coefficient de temps (time-weighted return - TWR).

L'approche MWR mesure le taux de croissance composé de tous les fonds du compte (incluant les flux monétaires ou "cashflows") durant la période évaluée. Les flux ne contribuent évidemment que pour la période durant laquelle ils étaient présents dans le compte. La formule 1.1 illustre le calcul selon cette approche (Bacon, 2008).

$$V_E = V_S \times (1+r)^Y + \sum_{t=1}^{t=T} C_t \times (1+r)^{Y-Y_t} \tag{1.1}$$

Avec :

V_S = valeur au marché au début de la période

V_E = valeur au marché à la fin de la période

r = rendement durant la période

Y = nombre d'unités de temps durant la période évaluée

C_t = t ième flux monétaire

T = nombre de flux monétaires

Y_t = nombre d'unités de temps depuis le début de la période au moment du t ième flux

L'approche TWR, à l'inverse, accorde le même poids à chaque période, peu importe le montant investi. Le rendement doit être calculé individuellement pour chaque période où un flux monétaire est enregistré. La formule 1.2 illustre le calcul selon cette approche.

$$\frac{V_1 - C_1}{V_S} \times \frac{V_2 - C_2}{V_1} \times \frac{V_3 - C_3}{V_2} \times ... \times \frac{V_{n-1} - C_{n-1}}{V_{n-2}} \times \frac{V_n - C_n}{V_{n-1}} = 1 + r \qquad (1.2)$$

Avec :

$V_t =$ valeur au marché juste après le flux monétaire C_t à la fin de la période t

Tel qu'indiqué par Lawton et Jankowski (2009), MWR est sensible à la taille et au moment où les flux monétaires sont effectués durant la période d'évaluation, alors que TWR ne l'est pas. Il s'agit d'ailleurs de l'une des raisons pourquoi TWR est préféré dans l'industrie. Selon Lawton et Jankowski (2009), cette approche mesure en quelque sorte les performances du gestionnaire de portefeuille, ajusté pour les flux monétaires, alors que MWR mesure les performances des actifs investis par le client, incluant l'impact des flux monétaires. Le gestionnaire n'ayant aucun pouvoir sur les flux dans le compte, il préfère logiquement les écarter de ses calculs pour ne pas être affecté positivement ou négativement par ceux-ci. TWR a été utilisé dans le cadre des projets de cette thèse.

La deuxième étape du processus d'analyse des performances, l'attribution de performances, vise à identifier les sources d'écarts entre les rendements du portefeuille et ceux d'une référence (benchmark) (Lawton et Jankowski, 2009). Considéré hors contexte, il est très difficile d'émettre un jugement sur un rendement. Un portefeuille peut afficher un rendement de -5%, s'il s'agit d'une période de crise économique et que le marché a perdu plus de 15% durant la même période, le portefeuille a plutôt bien performé ! L'attribution de performances tente d'expliquer à l'aide de différents facteurs cette différence de 10%. Il existe une multitude de modèles d'attribution comportant plus ou moins de paramètres selon le type d'actifs évalués et

le niveau de précision et de complexité recherché. Un modèle d'attribution est présenté plus en détail dans le chapitre 4.

La dernière étape du processus, l'évaluation des performances ("performance appraisal") tente de déterminer si les performances observées sont le résultat de la compétence du conseiller ou simplement de la chance, et si le conseiller risque de maintenir ce même niveau. L'évaluation des performances dépasse le cadre des projets de cette thèse.

Le thème de l'analyse des performances financières est commun aux trois projets présentés dans cette thèse. Le chapitre 2 présente un outil novateur de sélection permettant de facilement identifier et analyser des courbes représentant des rendements de portefeuilles ou des titres financiers. Le chapitre 3 tente de faciliter l'évaluation des performances en affichant un aperçu de la composition d'indices financiers sans modifier le contexte d'analyse, tout en limitant l'occlusion sur le graphique. La technique permet également de comparer plusieurs groupements de titres (ex : titres de secteurs financiers) de façon plus efficace que l'état de l'art. Enfin, le chapitre 4 présente une plateforme complète d'attribution des performances financières, un thème qui ne semble jamais avoir été traité par le passé par la communauté en visualisation.

CHAPITRE 2

VECTORLENS : ANGULAR SELECTION OF CURVES WITHIN 2D DENSE VISUALIZATIONS

Maxime DUMAS[1,2], Michael J. McGUFFIN[1], Patrick CHASSÉ[2]

[1] Département de Génie Mécanique, École de Technologie Supérieure,

1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

[2] Croesus Finansoft, 600 Armand-Frappier, Laval, Québec, Canada H7V 4B4

2.1 Abstract

We investigate the selection of curves within a 2D visualization by specifying their angle or slope. Such angular selection has applications in parallel coordinates, time series visualizations, spatio-temporal movement data, etc. Our interaction technique specifies a region of interest in the visualization (with a position and diameter), a direction, and an angular tolerance, all with a single drag. We experimentally compared this angular selection technique with other techniques for selecting curves, and found that angular selection resulted in a higher number of trials that were successful on the first attempt and fewer incorrectly selected curves, and was also subjectively preferred by participants. We then present the design of a popup lens widget, called the VectorLens, that allows for easy angular selection and also allows the user to perform additional filtering operations based on type of curve. Multiple VectorLens widgets can also be instantiated to combine the results of their filtering operations with boolean operators.

Keywords : Information Visualization, Finance Visualization, Interaction Technique, Selection Technique, Curves Selection

2.2 Introduction

Densely overlapping curves, trajectories, and edges occur in many visualizations, including parallel coordinates, time series data, GPS data showing movements of people or vehicles over time, node-link diagrams of graphs, etc. A fundamental task to support in such visualizations is the selection of subsets of curves. Currently, this is possible with a variety of techniques : clicking directly on curves, dragging out a region (by dragging a brush or by drawing a rectangle) that overlaps the curves to select, selecting curves according to some criterion such as "category" or "type" of curve, or sketching the shape of curves to select (Buono et Simeone, 2008; Holz et Feiner, 2009).

Another approach which has been less investigated to date is selection by specifying the angle or slope of curves. Such angular selection could be useful in many cases : selecting one of several overlapping clusters within a parallel coordinates plot, or one of several overlapping edge bundles (Holten et van Wijk, 2009), or selecting all the stocks rising or falling at a given rate on a given date, or all the boats moving along similar trajectories through a given region of an ocean. In some of these examples, automatic clustering might help identify groups of related curves, but clustering algorithms typically require tuning to decide how many clusters to find, whereas a user may visually identify very quickly a group of curves following some common direction and wish to select them.

In our work, we investigate a technique performed with a single drag for specifying (1) a region of interest within the visualization, (2) an angle or direction, and (3) an angular tolerance (Figure 2.1). This kind of interaction is well suited for integration into a lens widget that offers the user additional functionality, such as filtering options. As discussed later, our prototype interface allows a user to instantiate a lens widget, called a VectorLens, and select curves within it based on angular criteria. This is in contrast to previous techniques (Hauser *et al.*, 2002) that don't involve a lens and simply select a range of slopes over some horizontal interval, or other previous work (Hochheiser et Shneiderman, 2004) where a user paints multiple regions to define a compound selection of only curves passing through all such regions. Because our
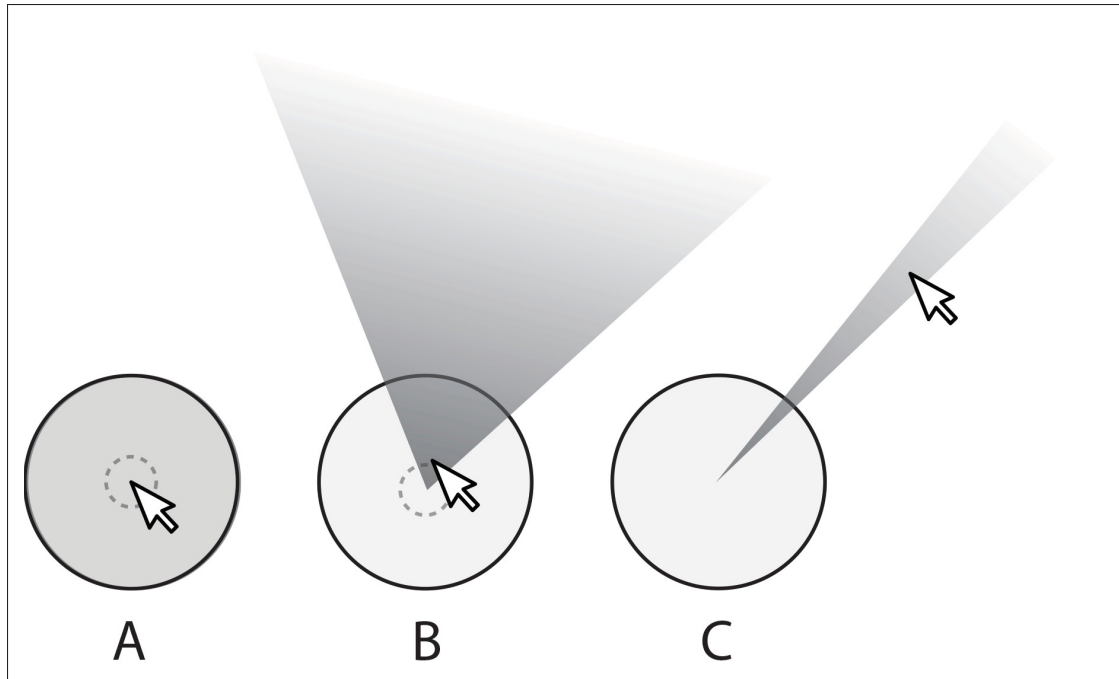
Figure 2.1   All curves passing through the outer circle are selected when the user stays inside the inner circle (A). Angular selection is enabled when the user moves outside the inner circle (B), and can be refined by extending further outside the circle (C).

approach is oriented toward use with a lens widget, we designed our interaction technique to enable the user to precisely select a subset of curves within the lens with a single click-and-drag, without having to instantiate a second lens (although the user may optionally do so, as presented later). Our approach could sometimes have an advantage over techniques that select slopes over a horizontal interval (Hauser *et al.*, 2002) because the selected curves are limited to a circular region in addition to a limited set of directions or slopes, making it more restrictive.

Before presenting our VectorLens widget, we first study the basic selection technique by which the user selects a region, a direction, and an angular tolerance, by comparing it with other curve selection techniques. Four techniques were experimentally compared : rectangle selection, a circular brush, single-drag angular selection (Figure 2.1, B and C), and a hybrid combination of circular brush with angular selection (Figure 2.1, A, B and C). In our experimental task, participants had to select a highlighted target subset of curves by using one of the techniques,

and could perform one or multiple interactions (for example, dragging out two or three rectangles in a single trial to select the subset of curves passing through all rectangles). Although the angular and hybrid techniques required slightly more time on average to complete the task, these two techniques also resulted in the largest number of trials that could be completed with a single drag, and also minimized the number of erroneous curves selected during a trial. (This is important for using the technique within a lens widget, as it greatly reduces the need for a user to create multiple lenses for compound selections of curves.) Participants subjectively preferred the hybrid technique overall.

Later, we present the design of the interactive VectorLens widget, which allows the user to perform angular selections in a single drag, as well as brush selections in a single click. Once invoked, the lens remains displayed on the screen, affording further interactions to move the lens, as well as filtering and category-based selection.

Our contributions are (1) the design of a single-drag angular selection technique and of a hybrid technique combining angular selection and a circular brush, (2) an experimental comparison of selection techniques, and (3) the design of the VectorLens widget integrating our hybrid selection technique with extensible filtering and category-based selection features.

2.3   Related Work

Techniques to aid in the understanding of dense visualizations of points or of curves include excentric labeling (Fekete et Plaisant, 1999), lenses that magnify content (Pietriga *et al.*, 2010), and lenses that filter out a subset of content (Ellis et Dix, 2006). In the case of curves, additional approaches for aiding visualization involve the interactive deformation or "bending" of curves (Henry Riche *et al.*, 2012).

Techniques for the *selection* of elements in a visualization can be used for invoking operations on the selected elements, and/or for changing rendering parameters (such as the color or alpha) of the selected elements. If the elements to select are small and localized (or point-like), selection of just one element can be aided by techniques that cause the mouse cursor to grow

or shrink based on mouse velocity and/or on the proximity of candidate targets near the cursor (Grossman et Balakrishnan, 2005; Chapuis *et al.*, 2009), making it easy for the user to select the nearest target. DynaSpot is one of the most promising of many proposed techniques for aiding such single-target selections. Other single-target selection techniques are notable for their use of direction : Splatter (Ramos *et al.*, 2006), Escape (Yatani *et al.*, 2008), and Click-and-Cross (Findlater *et al.*, 2010) all allow the user to first indicate a location of interest, where there may be multiple overlapping or nearby candidate targets, and then drag in a direction that has been assigned to only one target to complete the selection. In these techniques, the direction is not an intrinsic property of the candidate targets, but is rather assigned arbitrarily to disambiguate them.

Selection of *multiple* points or curves in a visualization can be performed with traditional techniques where the user "paints" a selection using a rectangular or circular brush. Our work focuses on the case where the user wishes to select multiple curves that follow a similar direction and that pass through a common region. Since these target curves may overlap with other curves going in different directions, traditional paint-based selection is ineffective unless the user can paint a *compound* selection of at least two regions, to select the curves passing through both regions. This approach is used with timeboxes (Hochheiser et Shneiderman, 2004), where the user can specify two rectangular regions over a visualization of time-series data to select all curves passing through both. Other techniques for selecting curves with a given direction, angle, or slope are angular brushing (Hauser *et al.*, 2002), "angular queries" (another technique proposed by Hochheiser et Shneiderman (2004)), and the operators of Guo *et al.* (2010) that have an "angular tolerance". The first two of these apply to parallel coordinates and time series data, but do not naturally apply in the case of parametric curves where neither *x* nor *y* is an independent variable, because the techniques assume there is some horizontal interval over which to operate, which would be meaningless in a parametric plot. Also, none of the three previous techniques (Hauser *et al.*, 2002; Hochheiser et Shneiderman, 2004; Guo *et al.*, 2010) is optimized to enable the definition of a selection in a single click-drag. In contrast, our present work proposes a technique that can be invoked in a single drag for fast execution, and can be

applied to parallel coordinates, time series data, or continuous parametric curves (Figure 2.3). Our proposed VectorLens widget is also more flexible and complete in its features than these previous widgets. For example, a single VectorLens widget allows the user to select multiple angular ranges without instantiating additional lenses.

Kosara (2011) proposes multitouch input techniques for selecting multiple curves of a given slope or angle using three or four fingers simultaneously. This can work well in some cases, however multitouch input suffers from a lack of spatial precision, as well as physical limits on how far and how close fingers can be positioned.

Various "query-by-sketch" techniques have been proposed (Wattenberg, 2001; Buono et Simeone, 2008; Muigg *et al.*, 2008; Holz et Feiner, 2009) to select curves by drawing the approximate shape of the curves of interest. These allow the user to be more specific in their selection than simply specifying a location and a direction, however "query-by-sketch" can also incur more cost to perform, since the user must provide the entire shape of a curve rather than simply a location and direction.

Later in our paper, we show how to incorporate angular selection with a single drag inside a more complete widget called the VectorLens, offering more options to the user. Tominski *et al.* (2014) recently published a survey of interactive lenses. Among them, an interesting widget for performing selections based on angles is the EdgeAnalyzer (Panagiotidis *et al.*, 2011), which detects clusters of edges, making it easier for the user to select these clusters. However, the EdgeAnalyzer does not allow the user to perform selections with a single drag as our VectorLens does.

In summary, our angular selection technique is unique in that it can be performed in a single drag ; and can be applied to parallel coordinates, time series, and visualizations of continuous curves ; and is also incorporated in a larger widget so that the user may optionally invoke it in a single drag or invoke the widget's additional options for more control, unlike all previous techniques.

2.4 Angular Selection in a Single Drag

Our selection technique leverages an elementary and useful characteristic of curves : direction. In many visualizations, curves that pass through similar positions with similar slopes are meaningfully related, e.g., they belong to a "cluster" or "bundle" or somehow represent similar behaviors. Our approach for angular selection allows the user to select such clusters or bundles (or subsets of them) in the absence of automatic clustering algorithms, which would in any case require fine-tuning by the user.

Prior to clicking, a circular brush follows the mouse cursor, remaining centered on the cursor. (This is the solid black circle in Figure 2.1.) When the user clicks down, the brush is fixed in place. If the user drags outside the dashed circle (Figure 2.1-B), this defines a vector originating at the brush center and pointing to the mouse cursor position (Figure 2.2). Only curves passing under the circular brush *and* with slopes oriented in approximately the same direction will be selected. The angular tolerance allowed (i.e., the angle covered by the shaded angular sector) decreases as the cursor is dragged further from the center of the brush (Figure 2.1). In our implementation, the angular tolerance is initially 70 degrees (when the mouse cursor is immediately outside the dashed circle), and decreases as $1/d$, where d is the distance (Figure 2.1-C). This inverse function allows a large value initially for coarse selection, but decreases quickly to enable precise selections without having to move too far from the lens center. Theoretically, this technique allows arbitrarily high precision as the distance increases.

Highlighting shows the user a preview of the curves that will be selected prior to releasing the mouse button.

The user may use the mouse wheel to change the diameter of the lens, allowing for the specification of a more or less precise location of interest. Such adjustment can be performed prior to clicking, or even after clicking and during dragging (in which case, the highlighted subset of curves is updated to reflect the lens' new diameter).
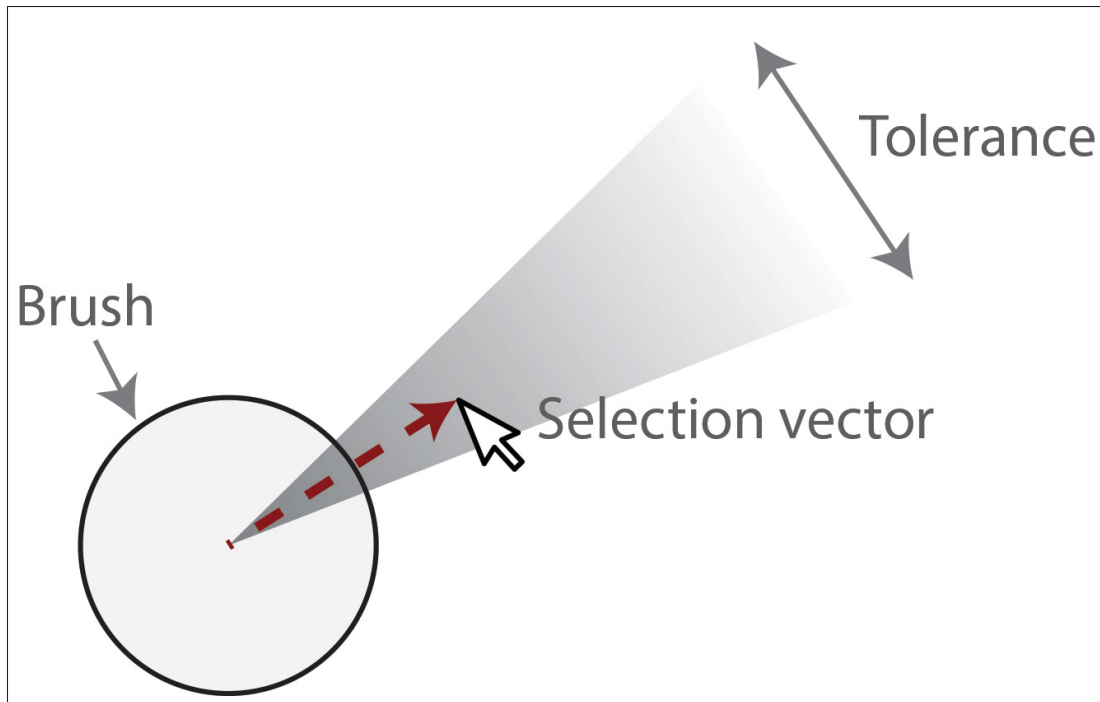
Figure 2.2   Elements of our angular selection technique. The brush specifies
a region of interest.

The slope of curves under the brush is calculated according to the direction of tangent vectors
at the intersections of the curves and the brush circle. If any of the tangent vectors' directions
fall within the range of directions of the shaded angular sector, then that tangent vectors' curve
will be selected (Figure 2.3).

An alternative approach would have been to compute the average slope of all tangent vectors
at the intersection points, or the average slope between the first and last intersection point, and
use this average slope to determine which curves to select. If the curve intersects the circular
region at two points, this is equivalent to using the *secant* slope. We did not implement this,
because such an average slope can be difficult for a user to predict (as well as being compu-
tationally more expensive), and it is easier for the user to drag in the direction of one of the
intersection tangents to select their desired curve(s). Furthermore, in practice, the "curves" in
parallel coordinates and time series plots usually have long straight-line segments, and would
not be as curvaceous as that shown in Figure 2.3 (right). If a user wanted to select a subset of
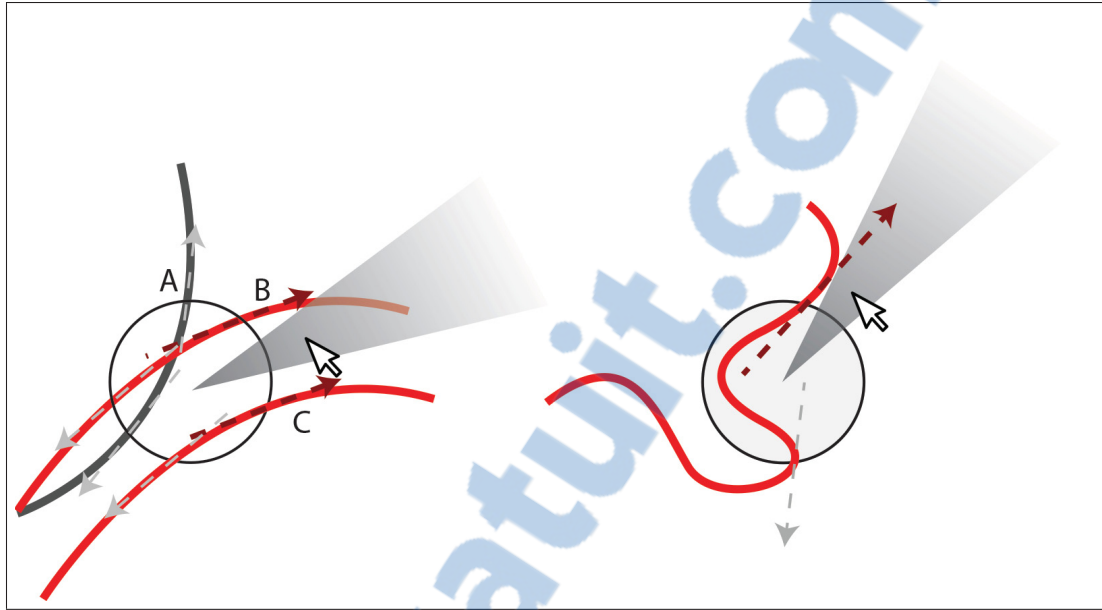
Figure 2.3   Tangent vectors are calculated at brush-curve intersection points.
Vectors in gray, such as A, are not selected as they do not fit in the angular range. Red
vectors such as B and C fit in the tolerance range, causing their curves to be selected.

curves in a parallel coordinates plot, they might first click on an axis and then drag out a range
of directions on the right side of the axis, for example, in which case they would not expect
their selection to be influenced by the slopes on the left side of the axis, which is another reason
to not use the secant or average slope.

Figure 2.3 shows that curves do not need to be *contained* within the angular sector to be se-
lected, but must have a tangent vector pointing in the same *direction*. The user can often easily
avoid unwanted neighboring curves by strategically positioning the brush prior to clicking, so
the brush does not cover unwanted curves. In extreme cases where the slope of the curve is
highly variable such as Figure 2.3 right, the user can simply solve this issue by reducing the
size of the lens or by leaving the lens centered on top of the curve.

By way of comparison, the "angular queries" proposed by Hochheiser et Shneiderman (2004)
limit the region of interest only along the 1-dimensional horizontal axis, Our technique allows
the region of interest to be positioned within 2-dimensional space.

The angular selection technique just described will be referred to as *Pure Angular* selection, which we evaluated experimentally. We also evaluated a *Hybrid Angular* technique that works just like the Pure Angular technique, except that the user may optionally release the mouse button within the 10-pixel-radius dashed circle in Figure 2.1-A, causing all curves passing through the brush to be selected. This shortcut allows the technique to be used as a traditional circular brush, which is faster than specifying an angle and is appropriate if there are no undesired curves under the brush.

Notice also the Pure and Hybrid angular selection technique can be applied to parallel coordinates, time series data, edges within network diagrams, and continuous curves.

2.5 Experimental Evaluation

To evaluate the performance of our angular selection technique, we experimentally compared it with other selection techniques in terms of speed, "errors" (our name for the number of curves whose selection state had to be corrected before the end of a successful trial), and subjective user preferences.

Our angular selection technique was chosen to be easy to integrate into a lens widget, and is designed to allow for selection of an angular subset within a circular region. In practice, this selection can often be done with a single click-and-drag, and with only one circular region defined. However, this click-drag may be somewhat slow to execute because of the precision required on the part of the user to define the angular tolerance. In contrast, other selection techniques (Rectangle and Brush selection, described below) may often require two click-drag actions, but have the advantage that each of these click-drag actions are easy and fast to perform. We chose these techniques for comparison to better understand the tradeoffs between these two styles of interaction : the first based more on a precise selection within a single circular region, and the second based more on fast, multiple selections.

2.5.1 Task

Participants were asked to select a target cluster within a synthetic data set that was displayed in a seven axis parallel coordinates chart. This is almost equivalent to a discrete time series over 7 time steps, and therefore we do not think a separate evaluation with time series would yield new insights. Four techniques of selection were compared : *Rectangle* selection, where the user drags out the diagonal of a rectangle ; *Brush*, where the user clicks to paint with a circular brush (optionally dragging before releasing, to paint a larger selection) ; *Pure Angular*, shown in Figures 2.1 B, C ; and *Hybrid Angular*, shown in Figures 2.1 A, B, C, which can be used as a circular brush by clicking and releasing, thus combining the Brush and Pure Angular techniques. Only one technique was available during each trial, but could be invoked multiple times to perform corrective selections or deselections.

With three of the techniques (Brush, Pure Angular, and Hybrid Angular), the user could adjust the size of the brush with the mouse wheel, both before and during a selection.

At the start of each trial, one of the four techniques is active, and the user is shown the full data set across 7 axes, with the target cluster of curves highlighted in red. The user then clicks and drags to activate the technique to attempt to select only the target curves. The successfully selected target curves are then highlighted in green, and the incorrectly selected non-target curves are highlighted in orange (Figure 2.4, top). If there are also target curves not yet selected, they remain highlighted in red. When the user releases the mouse button, if there are no incorrectly selected curves, the trial ends. Otherwise, the user must perform additional corrective selections (or deselections) with subsequent drags. The trial only ends when all the targets curves have been selected, and no incorrect curves are selected. The *Time* measured is the total duration of the trial. Furthermore, after each drag, we count the number of non-target curves that are erroneously selected, and the number of target curves that are not yet selected, and the sum of these numbers over all drags of the trial gives the total number of *Errors* for the trial. These are not errors in common sense of the term, because the user is forced to correct these incorrect selections so that there are none left by the end of the trial. However, we are interested

in counting such incorrect selections, to compare how close the user can get to the final target selection after just one click-drag, and for simplicity we refer to this count of incorrect selections over the course of the trial as Errors. In practice, techniques with smaller Error count may help the user deal with large datasets, because such techniques cause a smaller subset of curves to be highlighted after the first click-drag, possibly helping the user to more easily identify the subset of curves that interests them, and helping them understand what subsequent corrections (if any) are required.

Corrective selections and deselections can be performed in several ways. Clicking and dragging with the left mouse button adds additional curves to the previous selection (corresponding to a set union of the previous and new selections). Clicking and dragging with the right mouse button removes curves (corresponding to set difference). Holding down the space bar and clicking and dragging with the left mouse button selects *within* the previous selection (corresponding to set intersection). Set intersection is used in Figure 2.4, top and middle, where a 2nd rectangle is drawn to select only curves passing through both rectangles (in this case, corresponding precisely to the target curves, completing the trial).

Participants were free to use any combination of selections they liked, but were instructed to complete the trials as quickly as possible. We observed during pilots that set intersection with the spacebar was frequently the simplest and most effective approach to use, and encouraged participants to also use this approach when a single drag was insufficient to complete the trial. Usually, 2-3 drags were sufficient to complete a trial using intersection. For example, in Figure 2.4, top, a single drag with the Rectangle technique cannot select all target curves without also selecting erroneous curves (shown in orange), however a second Rectangle selection with set intersection completes the trial (Figure 2.4, middle). When using the Hybrid Angular technique, participants were encouraged to perform an initial angular drag for each trial, and if a corrective selection was needed, to do so using the brush shortcut with set intersection.

Participants were also advised to adjust the mouse wheel to obtain a brush diameter they were comfortable with, and then not frequently readjust it, as we found in pilots that frequently
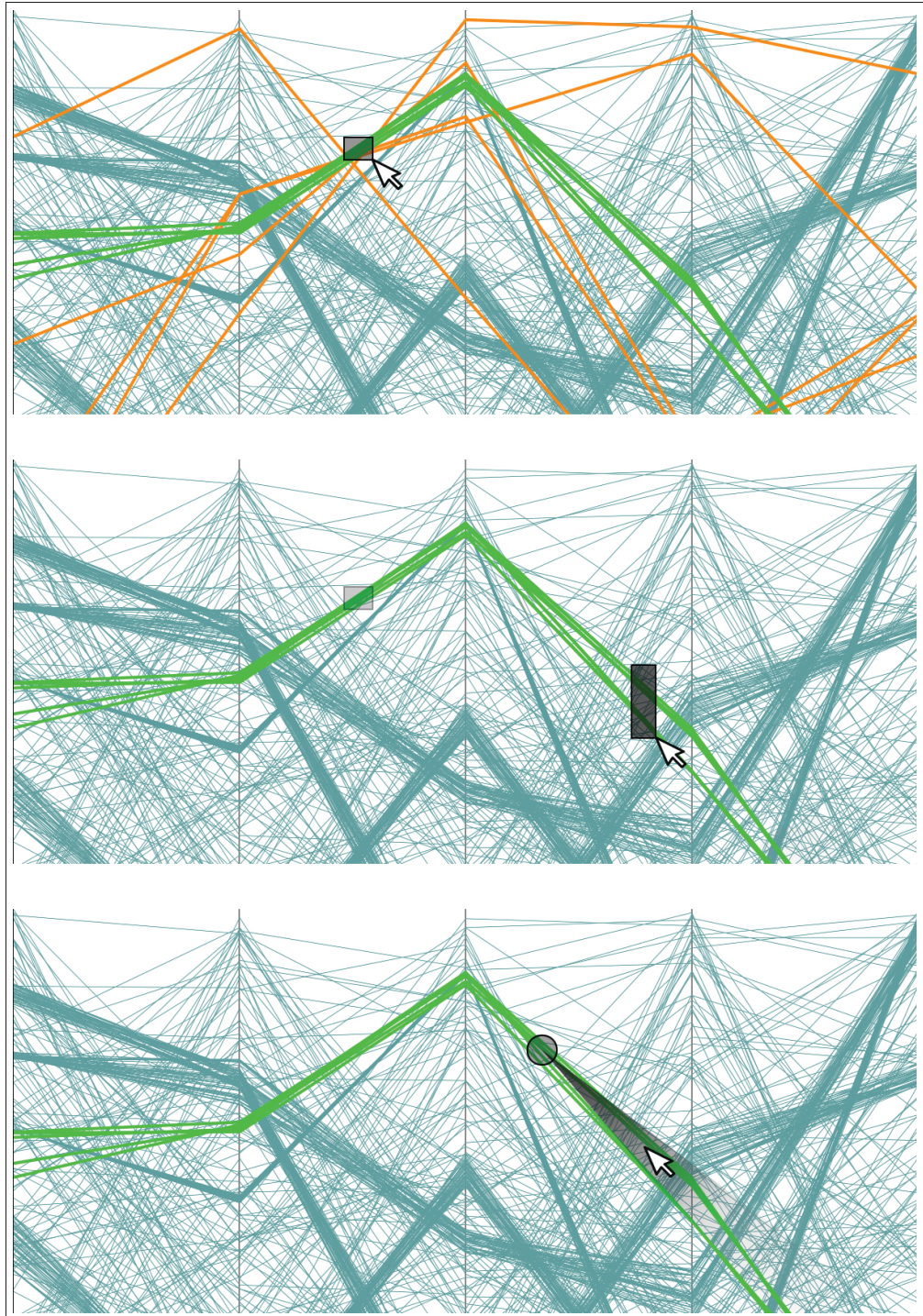
Figure 2.4   The rectangular selection requires two operations to obtain
the expected selection (top), whereas angular selection can select
the curves in a single operation (bottom).

readjusting the brush diameter was expensive in time. During the experiment, the selected brush diameter was persistently remembered across trials by the software, reducing the need to readjust it.

2.5.2 Data Sets

A new data set was randomly generated for each trial, with different random seeds used for each user. We limited our experiment to straight edges since our use case was finance oriented.

Many solutions have been proposed in the litterature (Albuquerque *et al.*, 2011; Bremm *et al.*, 2012) to generate synthetic data sets. We developped our own algorithm to control the density level on the charts. Each data set contained approximately 50% of its curves in 5 clusters, and the other 50% of its curves as uncorrelated background "noise". Within each generated cluster, each curve had to pass within 5% of the axis length of the randomly chosen centroid on each axis for that cluster. The target cluster always consisted of 5 curves, whereas the number of curves in other clusters and in the background depended on the desired *density* of data.

We define density as the fraction of pixels being used to display curves, taking into account the Euclidean length of line segments, without taking into account overlap between curves. For example, if a single horizontal line of thickness 1 is displayed within a 100×100 window, there are 100 pixels used to display the line out of 10000 pixels in total, yielding a density of 100/10000 = 1%. On the other hand, if the line is diagonal, extending from the lower left corner of the window to the upper right, its length is $100\sqrt{2}$, its thickness is 1, and its total area is $1 \times 100\sqrt{2}$, yielding a density of 1.4%. The density of a visualization with multiple curves is simply the sum of the densities for each curve, without considering overlap. If the total density of a set of curves is 100%, and the curves never overlap, then theoretically all pixels should be covered by curves, however in practice overlapping curves will leave many pixels uncovered even at a nominal 100% density.

In our experiment, we varied the density over 5 levels (Figure 2.5), from 5% to 90%. Our data set generation algorithm incrementally adds curves until the desired density is reached,

resulting in an average of 20 curves for 5% density, up to 380 curves for 90% density. Higher densities result in more overlapping curves, making it more difficult for the non-angular selection techniques to select only the target curves in a single drag.
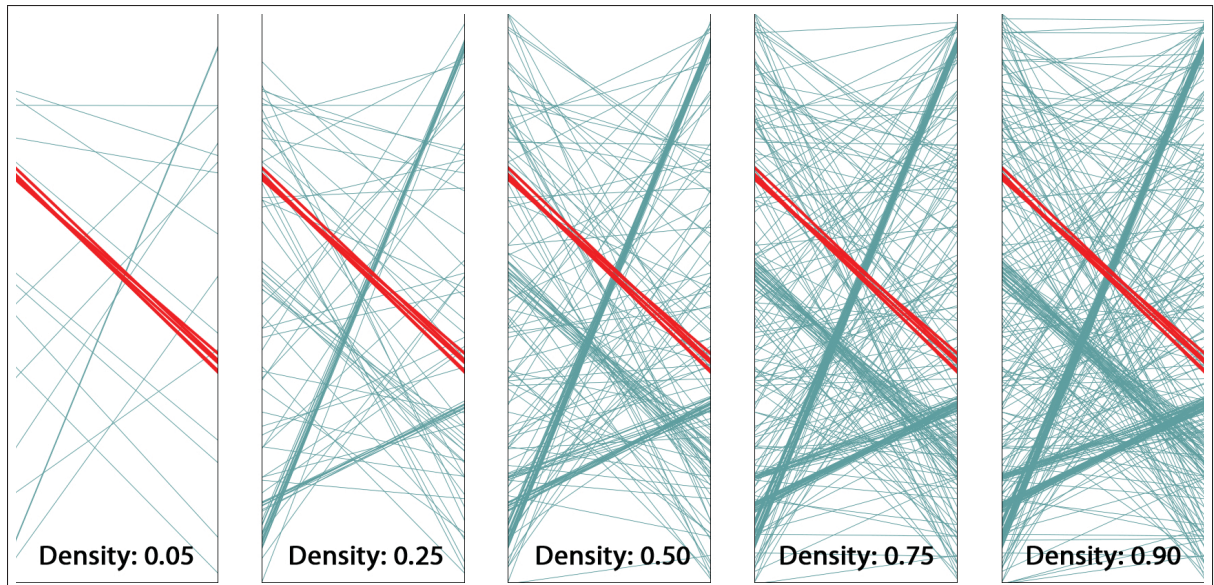
Figure 2.5   Cropped images showing the range of densities in the experiment. For all densities, there was a target cluster of 5 curves, initially highlighted in red.

2.5.3   Apparatus

The experiment was conducted on a 2.3 Ghz Quad Core Intel i7-2820 laptop running Microsoft Windows 7, connected to an external 24 inch $1920{\times}1080$ pixel LCD display. Participants used the external display, and external USB mouse and keyboard. Mouse acceleration was disabled.

2.5.4   Participants

To allow for counterbalancing of conditions and orderings, twelve volunteers (4 women, 8 men) participated, ranging in age from 23 to 46 (average 30.8, median 27.5). All were right-handed and controlled the mouse using their right hand. No one reported having physical handicaps, and none had color deficiencies. Nine participants were employees from Croesus Finansoft

working in different departments (programming, QA, business analysis, etc.). Three partici-
pants were master's students in engineering programs from ETS.

2.5.5 Design

The technique and density variables were within-subjects. Technique was countered-balanced
with a 4×4 Latin square. Each user performed 3 blocks of 50 trials with each technique, per-
forming first the 3 blocks for the first technique, then the 3 blocks for the 2nd technique, etc.
Pauses were allowed between blocks. The order of densities within each block was random.

Prior to the 3 blocks for each technique, participants were shown a video demonstration of the
technique, and then were asked to perform 50 warm-up trials (10 repetitions × 5 densities).

Not counting warm-up trials, there were

12 participants

× 4 techniques (Rectangle, Brush, Pure Angular, Hybrid Angular)

× 3 blocks

× 5 densities (5%, 25%, 50%, 75%, and 90%)

× 10 repetitions

= 7200 trials in total.

Each user session lasted approximately 1.5 hours.

2.5.6 Results

Figures 2.6 and 2.7 show the average time and errors for each technique and density.

Times were log-transformed and analyzed with a 2-way, 3 technique × 5 density ANOVA with
sphericity corrections. Both technique ($F = 29.2$, $p < 0.001$) and density ($F = 384$, $p < 0.001$)
had significant effects on time. Pairwise t-test comparisons with Bonferroni correction indi-
cates that all techniques are significantly different except for Rectangle and Brush. Rectangle
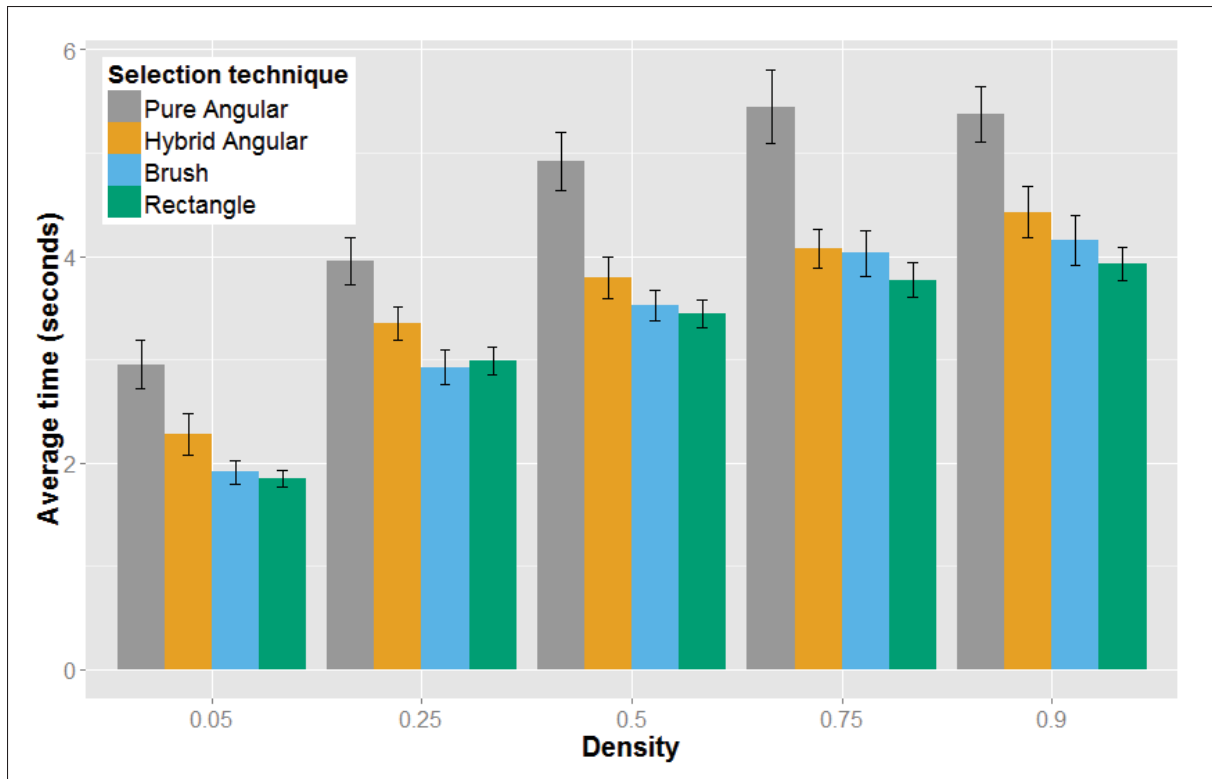
Figure 2.6   Average time per density level.

and Brush were the overall fastest techniques, with Rectangle slightly (but not significantly) faster.

Errors were also analyzed with a 2-way, 3 technique $\times$ 5 density ANOVA, again finding significant effects due to technique ($F = 25.3$, $p < 0.001$) and due to density ($F = 583$, $p < 0.001$). Pairwise t-test comparisons with Bonferroni correction indicates that all techniques are significantly different, with the two angular techniques the overall best.

We also compared the number of trials in each technique that were successfully completed with a single drag, i.e., completed with zero errors. The two angular techniques performed best by this criterion, as shown in the right-most column of this summary :

All tasks were divided in three blocks of 50 trials. After each block, the participant was allowed to take a pause. We compared the results between blocks and detected no significant difference for selection time and number of errors.
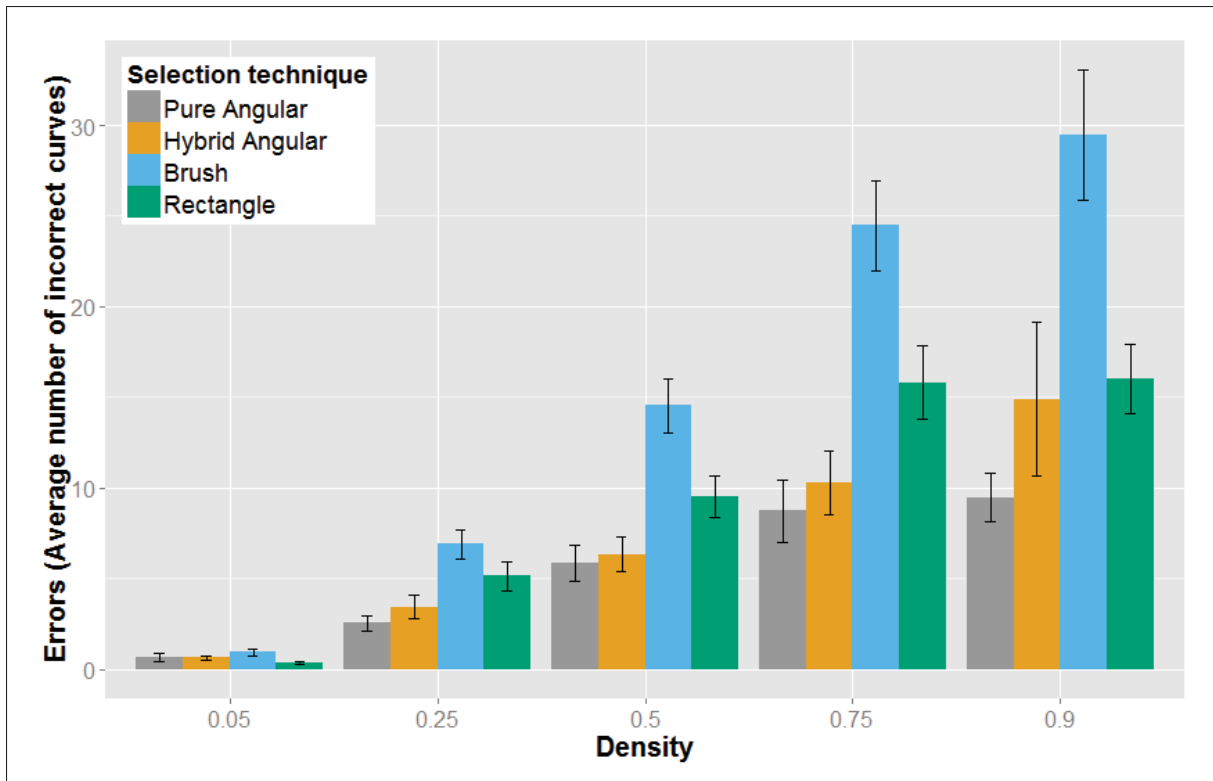
Figure 2.7   Average number of erroneously selected curves per density level.

Tableau 2.1   Detailed results

Technique	Avg. Time (seconds)	Avg. Errors	Number of single-drag trials
Pure Angular	4.53	5.45	522
Hybrid Angular	3.58	7.12	474
Brush	3.31	15.26	291
Rectangle	3.19	9.37	435

We compared the 4 techniques within each of the 12×5 (participant, density) combinations. We found that Hybrid Angular was the fastest of the four techniques in 8 out of the 60 such combinations. Density does not appear to advantage the hybrid as these 8 fastest times are

distributed similarly across all densities. No obvious pattern can be extracted from the data that would explain the ideal conditions for the hybrid technique to outperform the others speedwise.

Participants were observed to exhibit a variety of behaviors and strategies, resulting in varied performance. The fastest participants reported playing video games on a regular basis, with 3 of the top 4 fastest participants playing video games at least 15 hours per week. The two participants who were fastest overall, participants 2 and 9, were also the fastest at using the Hybrid Angular technique. Gaming performance is often influenced by the player speed at executing commands. Experience with advanced selection controls (such as the one found in video games) could have an impact on the observed results. Gamers clearly appeared to be more at ease with the angular controls, and rapidly developed strategies to optimize their selections. This might explain partly why they outperformed the other candidates using the hybrid approach.

After the experimental trials, participants were asked to choose their subjectively preferred technique. 9 out of the 12 participants preferred Hybrid Angular, and the remaining 3 participants preferred Rectangle. The most commonly given reasons for preferring the Hybrid Angular technique was that it allowed for greater "Precision / Limiting the number of selected curves" and allowed for "Rapid correction" of errors.

Because of the Latin square counterbalancing used in our experiment, 3 participants used the Hybrid Angular technique first, before trying the other techniques. All of these participants stated that they would have liked to re-do the trials with the Hybrid technique at the end, feeling that they would have done better a 2nd time.

2.5.7 Discussion

Of the four techniques, Rectangle is best in terms of time and better than Brush according to all criteria. The angular techniques are best in terms of errors and number of single-drag trials, and between the two, Hybrid Angular achieved a significantly better time than Pure Angular,

and Hybrid Angular was only about 12% slower than Rectangle. Hybrid Angular was also the most preferred subjectively.

We thus have a tradeoff between, on the one hand, Rectangle selection, which results in more erroneously selected curves after the first drag (e.g., Figure 2.4, top) but is nevertheless faster despite the extra time required for corrective selections, and on the other hand, Hybrid Angular selection, which is 12% slower on average but is more precise and was subjectively preferred.
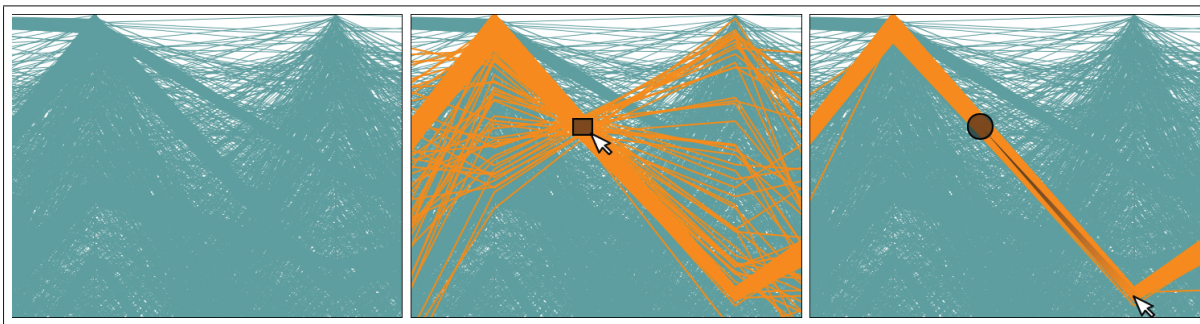
Figure 2.8   As the number of curves increases, a precise rectangular selection becomes almost impossible. Angular selection still performs very well to select the visible cluster, even at 1000% density.

We note again that our metric of "errors" does not refer to unsuccessful trials, since all trials ended successfully, but rather it refers to curves that were at some point in an incorrect selection state during the trial and required correction. Such (temporarily) incorrect selections are often unavoidable for techniques like Rectangle and Brush, and these techniques were nevertheless quite effective given their time performance, but the lower number of "errors" with the Angular techniques shows that the highlighted subset of selected curves is, on average, closer to the target subset after the first click of the user.

We suspect that one factor penalizing the angular techniques is that participants had extensive prior experience with performing rectangle selection in other software, but had almost no comparative experience with angular selection prior to the experience. A more longitudinal study might find that user performance with angular techniques improves significantly over time. This is also suggested by the wide variety of behaviors and results seen across participants.

Furthermore, in realistic scenarios where target clusters are not highlighted and densities are very high (Figure 2.8), the user may use these selection techniques to exploratively brush the curves during a drag, without necessarily knowing what direction a cluster has due to occlusion. In such a scenario, rectangle selection could be less effective than angular selection, since highlighting will only appear after clicking, and undesirable curves often cannot be deselected without restarting. In contrast, the use of angular selection allows the direction, tolerance, and circular brush size to all be adjusted during the first drag.

Finally, because angular selection more often allows selection to be completed with a single drag, this makes angular selection more amenable to inclusion within a lens widget with enhanced filtering options, as presented in the next section. A similar widget based on rectangular selection would be more likely to require multiple instantiations to complete a selection, leading to increased screen clutter, especially if the widget contains many auxiliary menus or options.

2.6 The VectorLens Widget

We propose incorporating angular selection within a lens widget called VectorLens. This widget subsumes the Hybrid Angular technique in our experiment, supporting angular and brush selection each in a single click, as well as supporting direct selection (by clicking on individual curves), categorical filtering capabilities, and complex queries involving multiple lenses. The VectorLens is designed specifically with times series and parallel coordinates in mind. Figures 2.9-2.10 show it applied to financial data.

When not inside a VectorLens, a dashed circle follows the mouse cursor, to provide a preview of the size of the lens that will be created if the user clicks. This dashed circle is resized with the mouse scroll wheel. Clicking down instantiates a VectorLens, after which the user may release (to select all curves in the lens, like with Brush selection) or may drag outward and release (to perform angular selection (Figure 2.9-A)). The user may then access several additional options within the VectorLens which we describe below, or dismiss the lens with the "X" icon at the
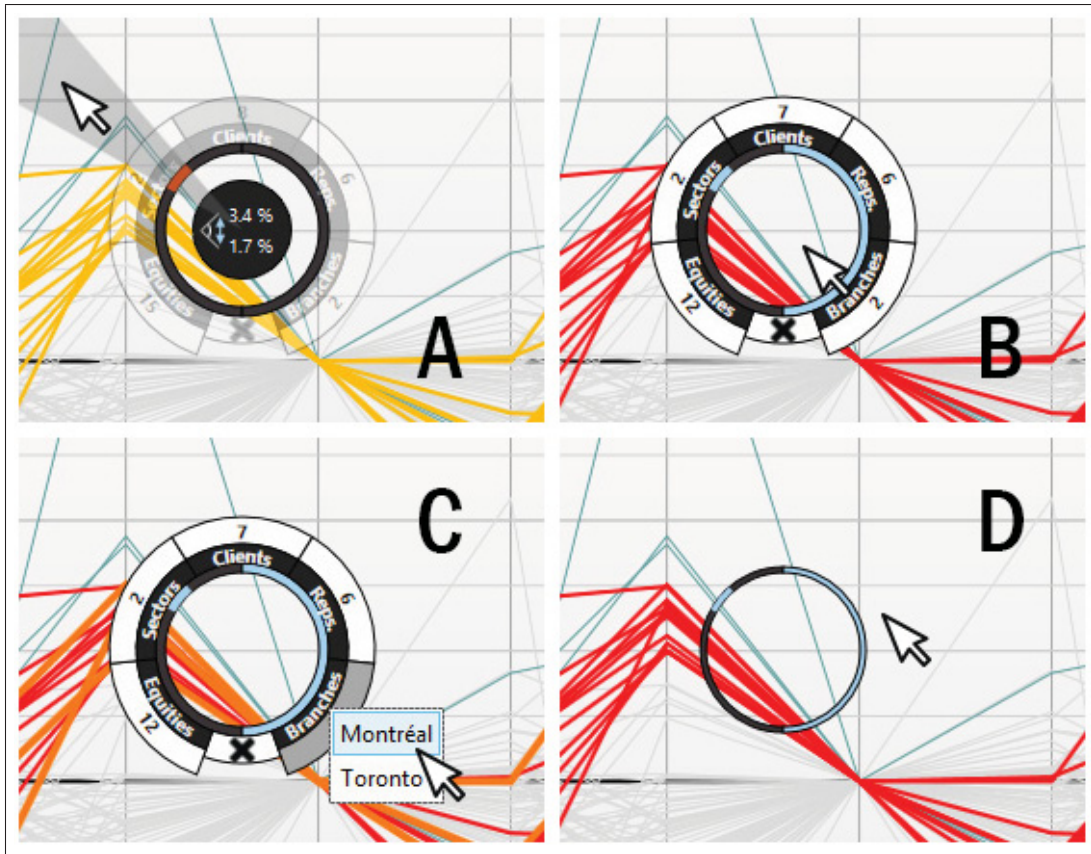
Figure 2.9   The VectorLens Widget.

bottom of the lens, or exit the lens in which case it compresses itself (Figure 2.9-D). After exiting, the user may instantiate more lenses in other places, or return to the first lens again to access its options.

When the cursor is inside an instantiated VectorLens, the user may click on the innermost ring to redefine an existing angular range, or add a new angular range, either on the left or right half of the ring. The user may also right-click on an angular range on the same innermost ring to remove a range. At any given time, there may be multiple non-overlapping angular ranges defined on the left and right halves of the VectorLens. For example, range R_1 on the left might correspond to curves between 25 and 45 degrees below the horizontal, range R_2 on the left might correspond to curves between 10 and 30 degrees above the horizontal, and range R_3 on the right might correspond to curves between 20 degrees below and 10 degrees above the

horizontal. The subset of curves selected in this case could be written as $(R_1 \cup R_2) \cap R_3$. In a time series visualization, the user might have instantiated a VectorLens on a discrete time instant, and in a parallel coordinates plot, the user might have instantiated a lens on a vertical axis. In both cases, the ability to select different ranges on the left and right is useful, since the polygonal curves under the lens will have one slope to the left, and typically a different slope to the right. For example, we might want to select different angles on both sides of an axis (entering in range A AND exiting in range B). Hence, the user may use the left slope, right-slope, or both, to discriminate and select curves. By default, multiple angle ranges on a single side are considered as disjunctions, and selection on both sides would be conjunctions.
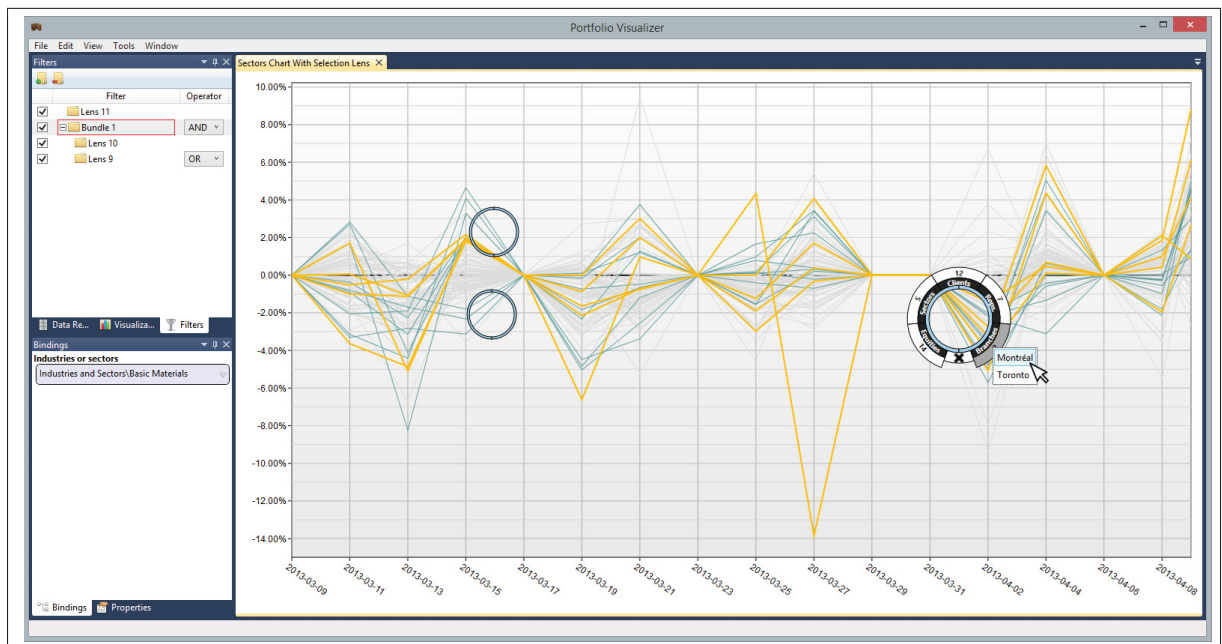
Figure 2.10   Complex query created using multiple lenses.

Other actions are also possible inside an instantiated VectorLens. The user may click-drag in the center of the lens to move the entire lens to a new location, making it behave like a magic lens (Bier *et al.*, 1993) for interactive filtering. The user may also position the mouse cursor directly over a curve in the middle of the lens (i.e., within a distance of 1 pixel of the curve) and click to select only that curve. The user may also click on the outermost ring to open a list of categories (Figure 2.9-C) and select one or more categories in the list, narrowing the selected curves to

only those that match the selected categories. This can be done for multiple criteria (displayed in the middle ring : "Equities", "Sectors", "Clients", "Reps", and "Branches" in the financial example in Figures 2.9-2.10), in which case the selected curves are narrowed to those matching the chosen categories within all criteria. Note that, prior to selecting within a criterion's list, the user may roll their cursor over the items in the list, to see the corresponding curves highlight (e.g., to see all curves corresponding to a particular Branch, for example). This allows the user to see which curves of each category are present, without having to display a large number of labels on the curves as could be necessary with Excentric labels (Fekete et Plaisant, 1999; Bertini *et al.*, 2009). The numbers displayed on the outermost ring show the number of categories available under each criterion. The user may also, at any time, use the mouse scroll wheel to change the diameter of the current VectorLens.

When multiple lenses are instantiated (Figure 2.10), by default an intersection of each lens' selection is computed, i.e., curves are selected if and only if they satisfy the constraints of all lenses. However, the user may modify this behavior using the query builder panel (left side of (Figure 2.10) to define groups of lenses and to choose the boolean operators to apply.

Normally, each VectorLens will only select among the curves passing through its circular lens area. However, an additional option changes this behavior, causing the widget to select among all curves passing through the horizontal interval covered by the VectorLens (Figure 2.11). This allows the VectorLens to imitate some previously described widgets, such asHauser *et al.* (2002). In this "vertical mode", the intersection points used to compute slopes of tangents are the intersections between curves and the vertical boundaries covering the lens' width. This feature gives more flexibility to the tool for special cases where the boundaries of the lens would be too limiting, allowing both constrained and unconstrained modes.

In summary, the VectorLens widget subsumes a more complete set of functions than the previously proposed widgets, while still allowing brush selection and angular selection to be performed each with a single click. It is worth noting that rectangle selection, which was the fastest technique in our evaluation, could also be incorporated into a rich lens-like widget, but
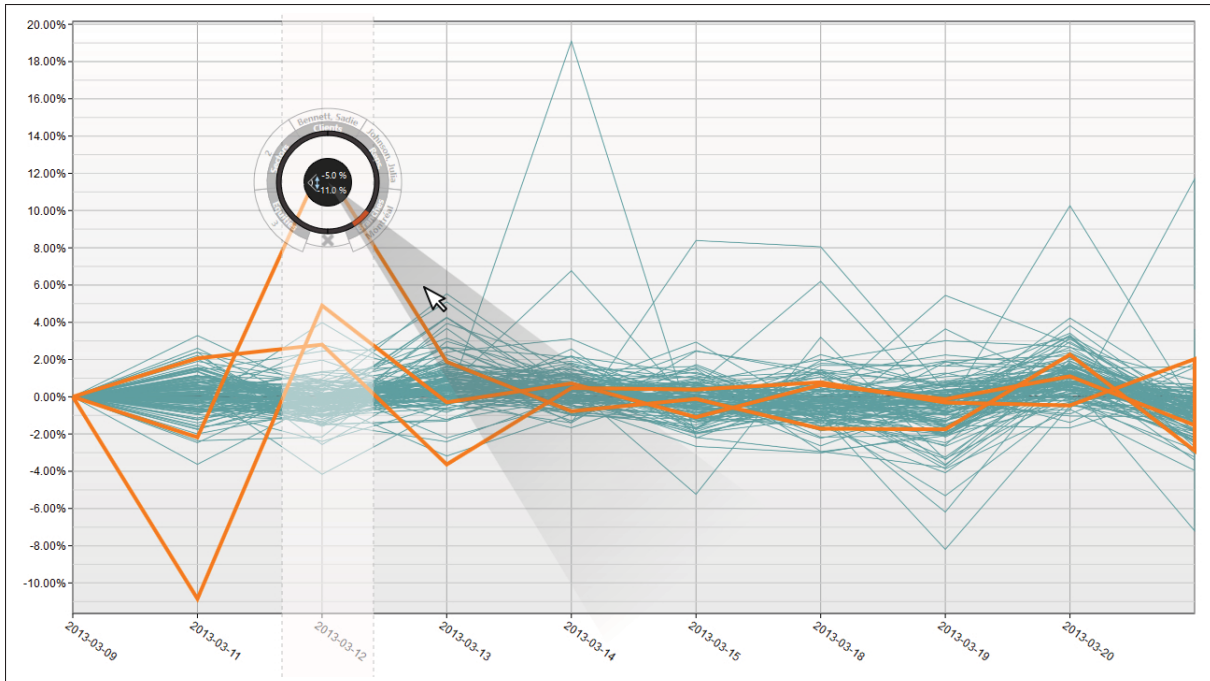
Figure 2.11   In this "vertical mode", the horizontal extent of the VectorLens is projected vertically, and all curves in this interval with the appropriate slope are selected.

this would likely require the user to instantiate more widgets (since selecting the same bundle of curves often requires two rectangle selections, compared with a single angular selection), leading to more screen clutter.

The fluid way in which a user can perform an angular selection and invoke a VectorLens widget in a single click-drag can be compared to other interactive techniques that are not concerned with the selection of data curves. For example, an extensive set of papers have proposed various menu techniques for selecting commands, many of these based on radial widgets (Kurtenbach et Buxton, 1993; Guimbretière et Winograd, 2000; Pook *et al.*, 2000). Many of these are surveyed in Bailly *et al.* (2007). FaST sliders (McGuffin *et al.*, 2002) allow a user to perform a rapid parameter adjustment in two drags, and then optionally keep a widget posted to the screen for further interactions, which is comparable to how a VectorLens is available for further interactions after its instantiation. None of these menuing techniques, however, are designed for the selection of elements in a visualization.

2.7　Conclusion

We have proposed two single-drag selection techniques for dense graphs : Pure Angular, and Hybrid Angular (which allows the user to perform angular selection or traditional brush selection). Both of these were compared in a controlled experiment with two other status quo single-drag selection techniques : rectangle selection, and brush selection. Our Hybrid Angular technique was slightly slower than the best of the four, but resulted in significantly fewer errors (i.e., fewer curves whose selection state required correction) over the course of each trial and more single-drag trials. Hybrid Angular selection was also preferred subjectively by 75% of participants. Finally, we presented the design of VectorLens, a selection widget that subsumes the Hybrid Angular technique without requiring any additional clicks, while also supporting multiple angular ranges, direct selection of single curves, selection based on one or more categories, and complex queries involving multiple lenses, making it more complete than any previously proposed widget for selecting curves.

Our angular selection tool was primarily designed to select clusters or bundles of curves close to each other and oriented in the same direction. Future work could explore automatic clustering techniques to make it easier for the user to "snap" their selection to the nearest cluster-like subset of curves. This could make it easier to eliminate the 2 outlier curves in Figure 2.8, for example.

Some participants commented during the experiment that our inverse function seemed too sensitive at first when adjusting the angular tolerance. Future work could therefore investigate multiple functions to find a better mapping from cursor distance to angular tolerance, which could improve the performance times with angular selection.

The VectorLens displays categories within simple linear lists. If the number of elements in such a list increased significantly, the user might lose time scrolling. Previous works, such as Ayatsuka *et al.* (1998) and Bederson (2000), provided interesting solutions to accommodate large quantities of data in lists. Such solutions could be applied to the VectorLens to improve scalability to large numbers of categories.

CHAPITRE 3

COMPRESSION AND SHIFTING OF LAYERS FOR REDUCING OCCLUSION IN MULTIPLE TIME SERIES

Maxime DUMAS[1,2], Michael J. McGUFFIN[1], Patrick CHASSÉ[2]

[1] Département de Génie Mécanique, École de Technologie Supérieure,

1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

[2] Croesus Finansoft, 600 Armand-Frappier, Laval, Québec, Canada H7V 4B4

3.1 Abstract

Visualization of multiple time series often suffers from overplotting, making it difficult to read the values of curves that are hidden by other curves. We present *compressed layers* and *shifted layers*, techniques for horizontally displacing the endpoints of line segments in discrete time series data that (1) enable the depiction of subsets of data and (2) reduce occlusion of endpoints, without sacrificing spatial resolution. Because endpoints are not displaced vertically, their vertical values can still be read and compared, but this comes at the cost of using a less familiar layout, and in the case of one variant of our techniques (namely, superposed compressed layers), it also makes the slopes of the line segments incomparable. We analyze design choices related the display of multiple layers (should they be superposed, or side-by-side, or offset ?), yielding three novel techniques : superposed compressed, juxtaposed compressed, and shifted. We also report an experimental comparison of three status quo techniques (conventional superposition, vertically stacked small multiples, and horizontally stacked small multiples) and our three proposed techniques over four tasks that were chosen based on real world tasks used in the analysis of financial data. Our results reveal tradeoffs between the six techniques, and show that our proposed techniques based on compression are competitive with status quo techniques.

3.2 Introduction

A common problem with time-series visualizations (Aigner *et al.*, 2007) is the occlusion of curves when visualizing many time-series (Figure 3.1A). Partitioning the data into multiple panes or windows (or "small multiples" (Tufte, 2001)) would reduce the amount of data shown in each pane, but would also reduce the amount of space available for each time-series, thereby decreasing spatial resolution and preventing both axes from being shared across all data curves.

Another way to depict distinct subsets of curves is to somehow display multiple *layers* of data in the visualization. In a 2D visualization of time-series, layers might be depicted using color coding (e.g., one subset of data curves in red, another in green), or using different curve thicknesses, or different levels of blur (one subset of curves blurred, another subset shown in sharp focus (Kosara *et al.*, 2001)), or by using movement (one subset of curves not moving, another subset vibrating (Bartram et Ware, 2002)). Each of these approaches, however, either does not reduce occlusion of data, or does reduce the spatial resolution available for the visualization, or both. There are also interactive techniques available for temporarily reducing occlusion and/or making layers more apparent, however these require a time investment on the part of the user to move a cursor or otherwise provide input.
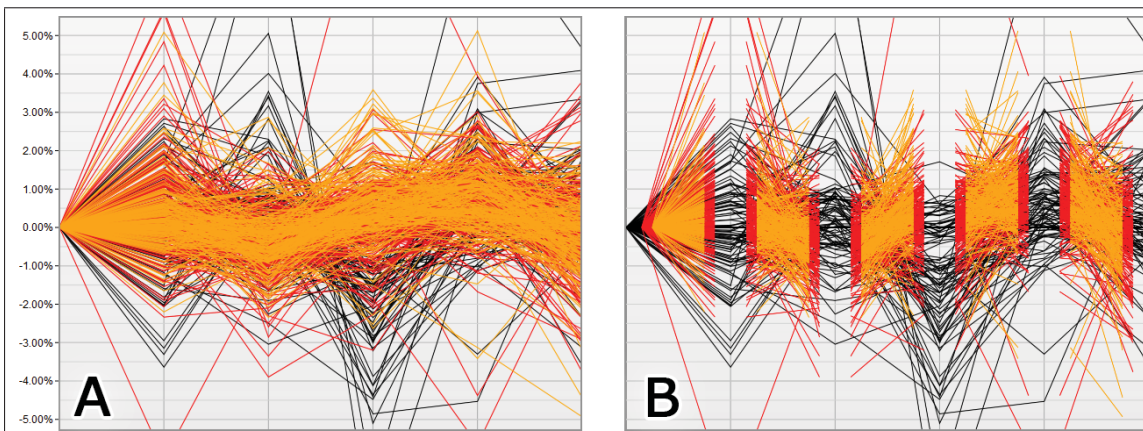
Figure 3.1   *A :* time series curves colored by type : black, red, or orange. *B :* the red and orange curves have been compressed horizontally to make 3 layers visible, and to eliminate occlusion of endpoints of line segments.

Jitter, i.e., the repositioning of points or endpoints of line segment by small (sometimes random) displacements, is a technique sometimes used to reduce occlusion in scatterplots (Figure 1 of Bezerianos *et al.* (2010)) and parallel coordinates (Figure 4 of Elmqvist et Tsigas (2008)). Jitter could plausibly be used in time series visualizations, to avoid overplotting where multiple time series achieve the same value. However, displacements in the vertical direction could make it impossible for the user to accurately read the values of time series data. Small *horizontal* displacements, however, would avoid such a problem.

We observe that in discrete time series, the space between consecutive time steps may be underrutilized, and investigate the use of horizontal offsets of different types to make subsets (layers) of curves more apparent. Figure 3.1B shows one of three approaches we have developed : To create a layer of curves appearing on top of lower layers, each interval is "compressed" horizontally, resulting in a (superposed) *compressed layer*. This compression changes the apparent slopes of the curve segments, but does not change the vertical positions of endpoints (meaning that values can still be compared within and between groups), and reduces occlusion with no loss in spatial resolution. Layer compression can also be used as a focus+context technique (Cockburn *et al.*, 2008), where the context is on a background layer, and the focus is on a "foreground" (compressed) layer, as shown on Figure 3.2. This technique enables the display of additional contextual information without creating occlusion of the original values near the axes.

As presented later, we also propose two other techniques that horizontally displace endpoints : juxtaposed compressed layers, and shifted layers. Figure 3.3 compares various techniques for separating subsets of curves, with our three proposed techniques appearing as the right-most columns. In this figure, every row is phrased as a positive quality, so that satisfying more criteria is more desirable. Since our techniques are the only ones to satisfy the first three criteria, and two of them even satisfy all four, further investigation is warranted. Note that our proposed compressed and shifted layers may not be best as a *replacement* for other techniques, but could be combined synergistically with other techniques, such as color (which we evaluate later) and/or interactive querying.
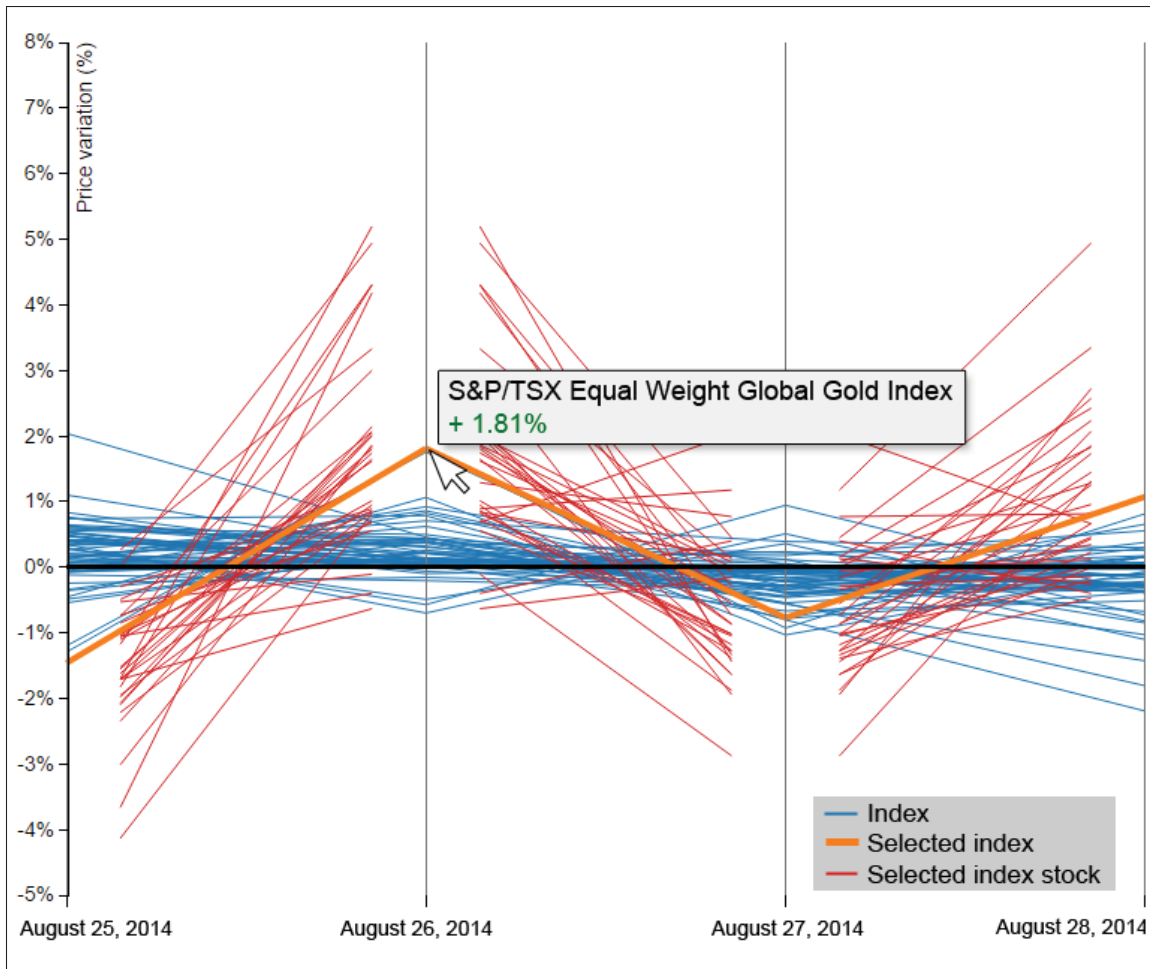
Figure 3.2   Compression used to display context on financial data. Blue lines show major Canadian S&P/TSX index performances for a four day range. The user selected a single index (in orange), for which the constituent companies are shown on an inner layer (in red).

In this work, we analyze design choices related to horizontal displacement of end points for the purpose of displaying multiple layers, yielding three novel techniques : two kinds of compressed layers (superposed and juxtaposed), and shifted layers. We also report an experimental evaluation of our three techniques with three status quo techniques (conventional, horizontally stacked small multiples, and vertically stacked small multiples), over four tasks. In all six techniques, layers were coded with a unique color. Our results demonstrate that there are tradeoffs between the techniques, and that our proposed techniques are competitive with the status quo. In each of Tasks 2 and 3, one of our techniques was best (as judged by average error), and in

Task 4, superposed was best (as judged by success rate). Over all tasks, our three techniques resulted in the lowest average errors.

	Color	Blur / Transparency	Movement	Random Jitter	Small Multiples	Layer Compression	Layer Shifting
Incurs no loss of spatial resolution	Yes	No	No	No	No	Yes	Yes
Reduces occlusion of endpoints of segments	No	No	Maybe	Yes	Yes	Yes	Yes
Enables inter-group comparisons of y-values (except where occluded)	Yes	Yes	Yes	No	Yes	Yes	Yes
Enables inter-group comparisons of slopes (except where occluded)	Yes	Yes	Yes	Yes	Yes	No	Yes

Figure 3.3   The three right-most columns show that our proposed techniques are the only ones exhibiting the advantages listed in the first three rows.

3.3   Related Work

Visualization of time series data (Aigner *et al.*, 2007) is a challenging topic, partly due to the long time spans involved, and due to the large number of curves that may need to be visualized, causing occlusion. Researchers have proposed collapsing curves to single rows of pixels (Kincaid et Lam, 2006) to eliminate occlusion, using interactive filtering (Hochheiser et Shneiderman, 2004) to reduce the number of curves to display, as well as other feature-rich interactive systems (Kincaid, 2010; Zhao *et al.*, 2011b,a). Our proposed technique is not interactive, and could be combined with ideas in these other works. Also, because we do not collapse curves into rows of pixels as in Kincaid et Lam (2006), the original shape of the curves remains visible.

Our work is based on the idea of partitioning a set of curves into subsets, and showing each subset on a different layer. Layers are a common theme in information visualization, and can be depicted in many ways. Layers can be embedded in a 3D space (Collins et Carpendale, 2007; Streit *et al.*, 2012) or can even be physically realized with special display equipment (Aboelsaadat et Balakrishnan, 2004). We seek an effective way of showing layers that will

make individual subsets of curves easier to read, without requiring additional interaction, or the complexity of 3D navigation, or specialized hardware.

Macroscope (Lieberman, 1994, 1997), transparent menus (Harrison *et al.*, 1995; Harrison et Vicente, 1996), Multiblending (Baudisch et Gutwin, 2004), and free-space transparency (Ishak et Feiner, 2004) are all ways of using transparency to show multiple images at the same time. This general idea could also be applied to time-series visualizations, but could be confusing when curves are dense. The use of transparency can also come at the cost of some spatial resolution, or creates many possible mixtures of alpha-blended colors that are difficult to distinguish (for example, 4 different original colors can produce at least $4^2 - 1 = 15$ different mixtures, not counting different orderings of colors).

Several interactive techniques can be used to more effectively browse through time-series data or other visualizations involving thin curves. These include user-initiated deformation (Wong *et al.*, 2003; Riche *et al.*, 2012), user-initiated vibration (Ramos *et al.*, 2006; Wong et Carpendale, 2007), interactive displacement of layers and with a 2.5D orthographic projection (Ramos *et al.*, 2006; Lopez-Hernandez *et al.*, 2010), lens-based techniques (Hochheiser et Shneiderman, 2004; Ellis et Dix, 2006; Pietriga *et al.*, 2010), as well as more general techniques like search and filtering. All of these can make visualizations more effective, but always at the cost of some investment in input on the part of the user. Because our proposed compressed layers have no required interactive techniques, they could be combined with many of these previous techniques to make them more powerful. The scope of our current work, however, is to study compressed layers in their static form.

The visualization of financial data has been a topic of growing interest, as shown by several recent works (Keim *et al.*, 2006; Schreck *et al.*, 2007; Ziegler *et al.*, 2010; Schaefer *et al.*, 2011; Dumas *et al.*, 2014), some of which have proposed techniques for aggregating curves, using small multiples, or using color coding to enable the comparison of a single time series curve with a market's performance. Our work is complementary to this previous work, as we propose a way of making multiple groups of curves visible simultaneously.

3.4   Data Layers

The partitioning of data into layers is particularly useful for tasks where the user must distinguish different types of curves. As shown in Figure 3.1A, the use of color is insufficient to clearly separate overlapping curves. Compression, for example, allows each group to be distinguished, with no occlusion of the endpoints of line segments.

3.4.1   Context

Our compressed and shifted layer techniques were designed with the intention of applying them to financial data. We identified several situations where datasets containing many financial time series must be compared. For example, it might be useful to compare the historical prices of securities between multiple industries, sectors, or asset classes to visually detect trends, anomalies, etc. Such comparisons might be used to to create a more diversified portfolio with uncorrelated assets from different markets. In our visualization, each layer might correspond to one industry or sector.

3.4.2   Compression vs Masking

Our technique leverages the horizontal space between consecutive time instants t_i and t_{i+1} to display layers. We assume that this intervening space contains no intermediate data, and therefore the consecutive points for t_i and t_{i+1} are connected by a straight line segment. This segment can be transformed, in theory, with no loss of information, as long as the user can still ascertain the original values. At least two transformations are possible that could allow the user to distinguish the values on each layer.

First, each line segment could be *compressed* (Figure 3.4A) horizontally. This compression changes the horizontal positions of the endpoints, and also changes the apparent slope. However, because it is the vertical position of each endpoint that shows its time-dependent value, these values can still be read with no occlusion.
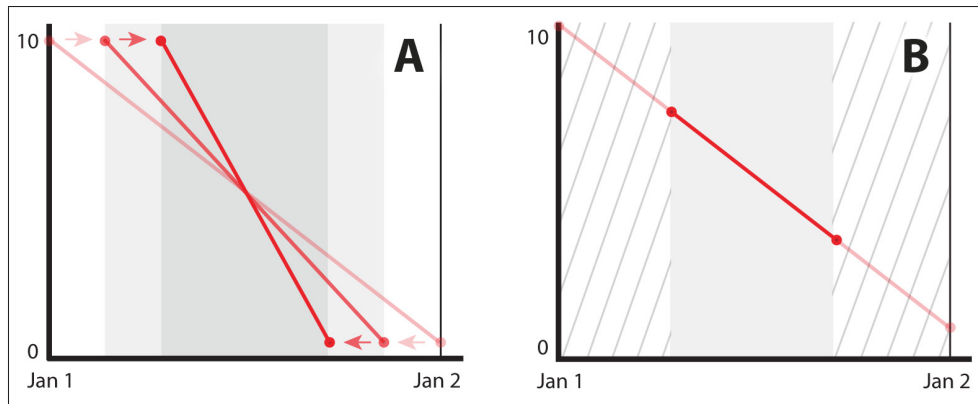
Figure 3.4   Two approaches for separating the line segments of time series curves into layers : (*A*) compression and (*B*) masking.

Second, each line segment could be *masked* (Figure 3.4B). This has the advantage of preserving the original slope, but erases the original endpoints, requiring the user to imagine where the the resulting line segment would intersect with each axis if the line segment were extended. Such masking is comparable to the effect of window blinds, picket fences, or bezels in tiled displays, and for certain tasks has been shown to not be detrimental (Bi *et al.*, 2010).

We implemented both the compressed and masked approaches for creating layers. Our informal testing found that masked time series were much more difficult to interpret, since it was typically unclear where the imaginary intersection points were, especially with line segments having a large slope, and especially with many layers present (causing incrementally more masking of the top-most layer).

We showed both approaches to 5 financial experts. With the compressed approach, all of them found the distortion of slopes to be initially somewhat confusing, but after an explanation were comfortable with it. In contrast, all 5 of them found the masked approach to be much too difficult to interpret.

We decided that it was more important to faithfully show the original vertical values of the time series than to preserve slopes, and therefore dropped the masked approach and retained the compression approach for further investigation and evaluation.

3.4.3 Layout Styles

Figures 3.1B and 3.5B show how multiple compressed layers can be **superposed**. Figure 3.5C shows that the upper layers could also be **juxtaposed**, and Figure 3.5D shows that layers can also be shifted to avoid occlusion.
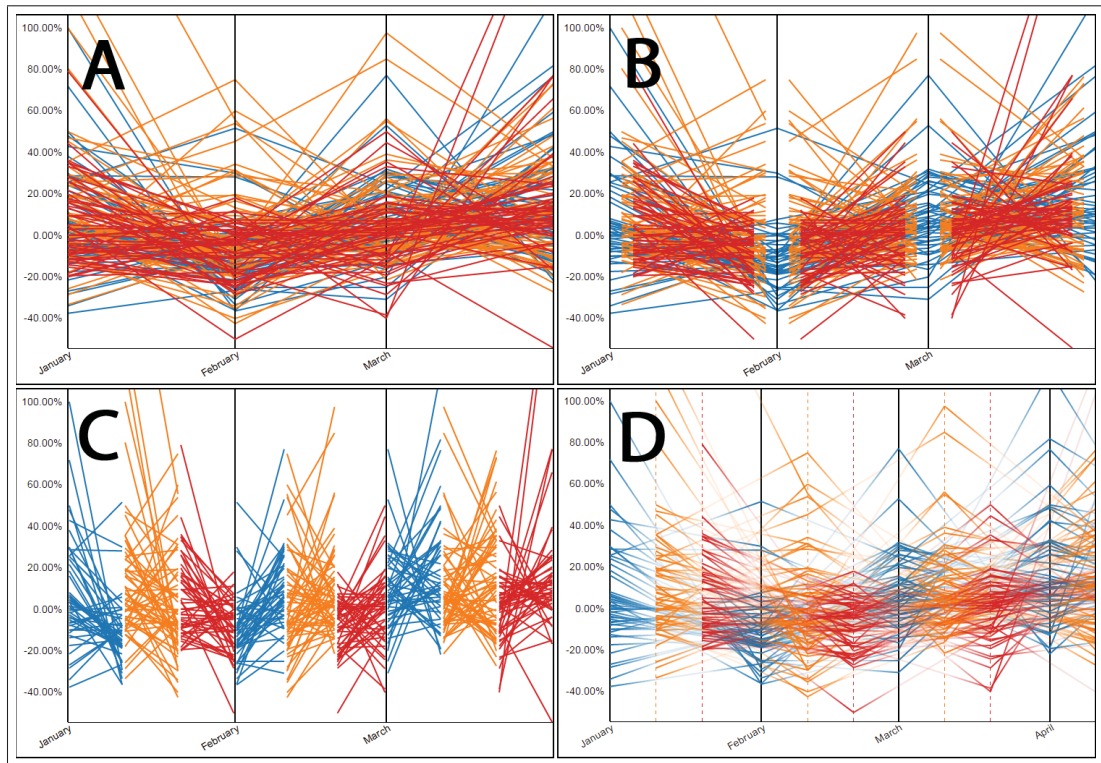
Figure 3.5 Layout styles for 3 layers (red, orange and blue). *A :* status quo with only color coding and no compression. *B :* superposition. *C :* juxtaposition of the upper 3 layers. *D :* shifted layers.

Superposition results in upper layers that undergo incrementally more compression, causing more and more distortion to the slopes of the line segments. In contrast, juxtaposition compresses all upper layers by the same amount, allowing slopes of upper layers to be more easily compared, but also compressing all the layers more than they would be if only superposition were used.

Shifting proposes to avoid end point occlusion without introducing distortion and discontinuity like compression does. It simply shifts upper layer inside the tick space by a predefined spacing. For example, in figure 3.5D, the first blue, two colored axes are added between each month for yellow and orange. The shifted ticks are always associated with the previous black tick (the previous month in this case). An opacity gradient is added between ticks to allow underlaying layers ticks to be visible. In our experiment, opacity was set to 10% between the ticks, and 100% at the tick.

Section 3.5 presents a controlled experiment that measured performance with these different layout styles, to gain an understanding of their limits and relative performance.

3.4.4 Limitations

Because we leverage unused space between consecutive time instants, our approach breaks down if the intervening space contains intermediate data, or (equivalently) the user has zoomed out so that many time instants are visible and there is little space between them. For example, Figure 3.6 shows data on 3 layers at 24-hour intervals. If the dataset over the same 3-day period contained details at 5-minute intervals which the user wanted to see, we could apply compressed layers with a 24-hour spacing, displaying the intraday information on each layer, and allow the user to interactively "raise" a given layer on mouse hover to make its details visible.

3.5 Experimental Study

We evaluated our compression and shifting techniques against state of the art methods for presenting layers : conventional (layers are simply encoded using colors) (figure 3.5A), horizontally-stacked small multiples (HSM) (figure 3.7A) and vertically-stacked small multiples (VSM) (figure 3.7B).

Conventional layers suffer from overplotting, sometimes making it impossible to perceive the locations of endpoints or even hiding entire curves. Small multiples, on the other hand, have
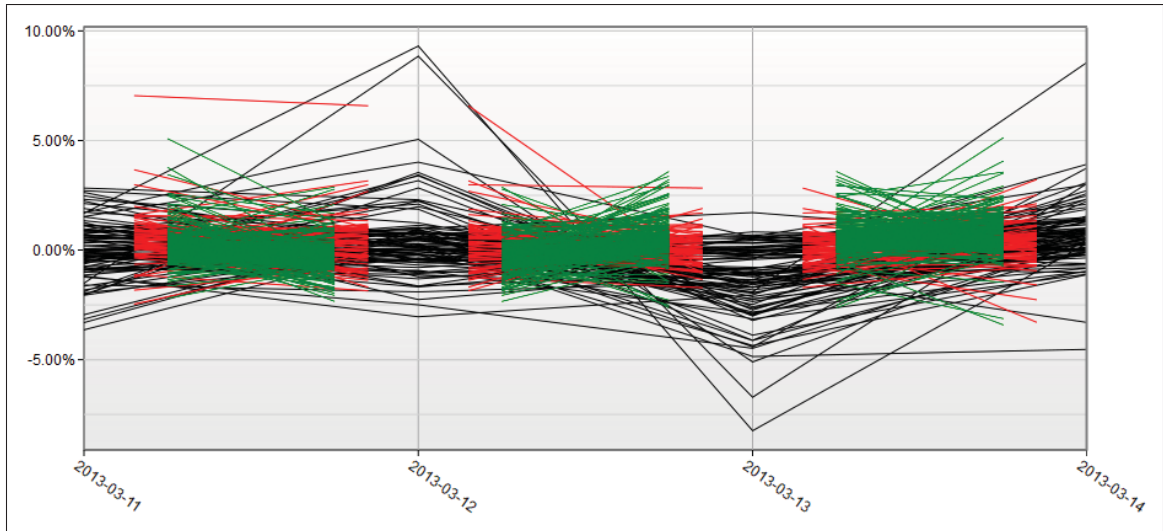
Figure 3.6   Daily changes in the values of stocks in 3 sectors of the NASDAQ, with each sector shown on a different layer.

the disadvantage of not sharing either $x-$ or $y-$axes, which might increase the amount of eye motions when comparing curves of different groups.

Our proposed techniques based on compression and shifting are not familiar to users, and may incur other costs. Furthermore, our superposed compressed layers technique modifies the scale of each layer by a different amount, making it difficult or impossible to compare slopes between groups.

3.5.1   Apparatus

The experiment was conducted on a 2.3 Ghz Quad Core Intel i7-2820 laptop with 8GB RAM and an NVIDIA Quadro 2000 GPU running Microsoft Windows 7, connected to an external 24 inch 1920×1080 pixel LCD display. Participants used the external display, as well as an external USB mouse and keyboard. Mouse acceleration was disabled.
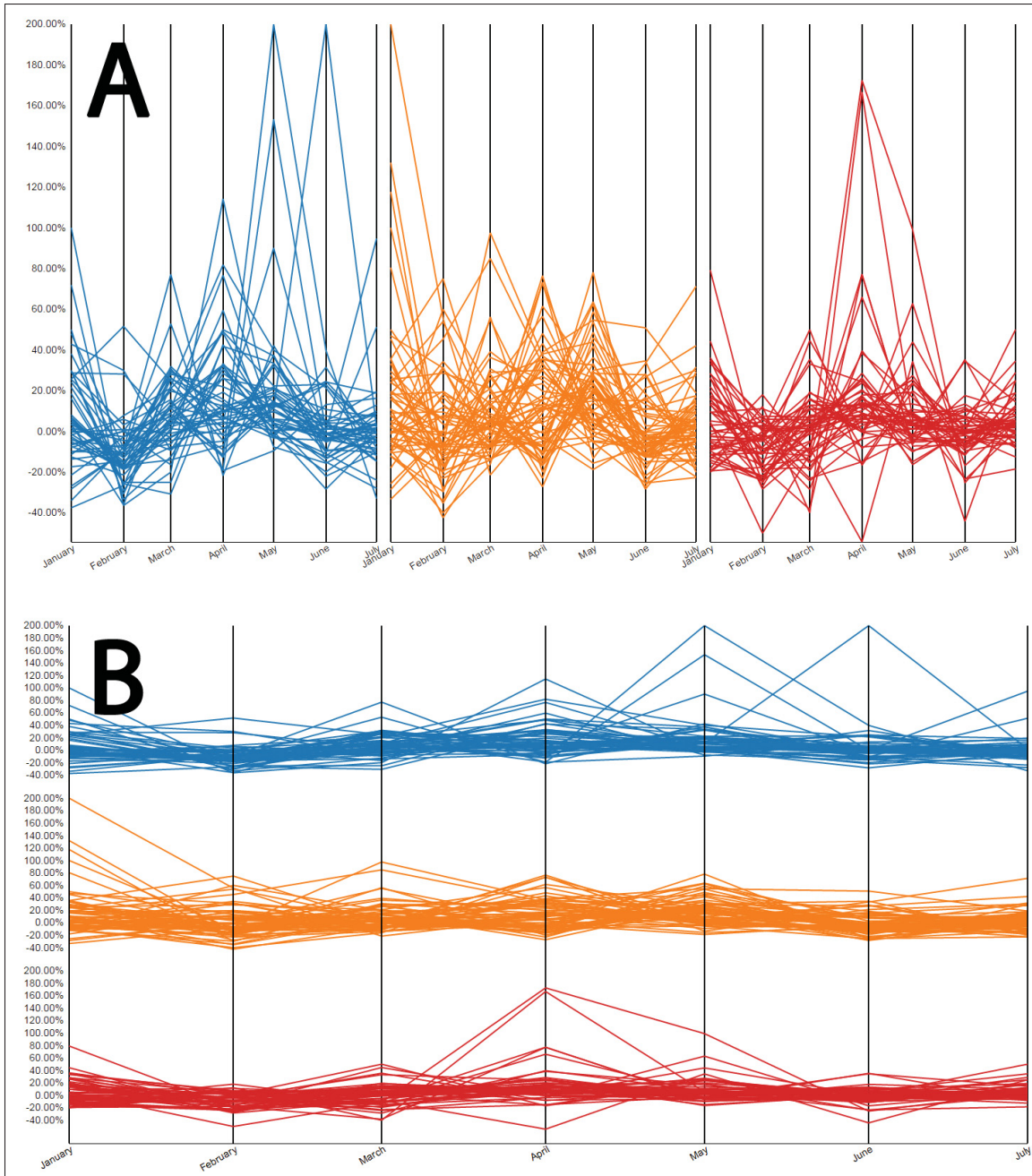
Figure 3.7 Compression techniques were compared to small multiples, where layers are positioned next to each other. *A :* Horizontal small multiples (HSM). *B :* Vertical small multiples (VSM).

3.5.2 Participants

Twelve volunteers (2 women, 10 men) participated, ranging in age from 19 to 43 (average 30.5, median 32). All were right-handed and controlled the mouse using their right hand. No one reported having physical handicaps, and none were colorblind. All participants were employees from Croesus Finansoft.

3.5.3 Design

To determine a set of tasks to evaluate the tradeoffs between these techniques, we chose a concrete application domain : financial data analysis.

We considered several possible tasks, involving comparisons of x values or y values, involving single values or ranges of values, involving slopes, and involving individual curves or groups of curves. This was narrowed down to the following four tasks which touch on all of these possibilities and which all have analogues in financial analysis. In all tasks, there were 3 groups (or 3 layers) of curves, and these were always color coded with the same 3 colors. Also, the $x-$axis had 12 tick marks labelled with the months of a year. We tried with different subperiods in a preevaluation and didn't find any interesting correlation based on the number of marks. We fixed this variable to limit the number of trials for the partipants.

Task 1 : compare y values of highlighted curves at a given x value. In this task, one curve in each group was highlighted (by drawing it thicker and with a white outline surrounding it), and the user was asked : *Of the highlighted curves, which one performed the best and the worst during the month of [month] ?* The user answered by examining the visualization and then choosing one of the three groups in a set of radio buttons.

Task 2 : compare and estimate range of y values of curves that are not highlighted at a given x value. In this task, no curves were highlighted, and the user was asked : *In which group do we find the curve that performed the best during the month of [month] ? Also, estimate the vertical range covered by the 3 top performing curves (one from each group) during that month.* For

example, if one group's top curve had a value of 100, another group's top curve had a value of 90, and another was 105, the range covered would be 105-90=15. The user answered by examining the visualization and then choosing one of the three groups in a set of radio buttons, and by entering the range in a text field.

Task 3 : compare ranges of *y* values at a given *x* value. In this task, no curves were highlighted, and the user was asked : *Which group exhibits the lowest volatility (smallest vertical range) during the month of [month] ? Estimate the vertical range covered by that group in that month.* The user answered by examining the visualization and then choosing one of the three groups in a set of radio buttons, and by entering the range in a text field.

Task 4 : compare slopes over a range of *x* values. In this task, no curves were highlighted, and the user was asked : *Which group seems to have performed the best over the time range from [month1] to [month2] ?* This required the user to look at each group as an aggregate of curves, and estimate the average slope of the group over the given time span. The user answered by examining the visualization and then choosing one of the three groups in a set of radio buttons.

In these tasks, the "groups" can be thought of as representing financial sectors, such as health care, technology, and energy ; and the curves can be thought of as equities such as stocks. Tasks 1 through 3 displayed curves showing the percentage of change in value from one month to the next for each curve. These percentages could be positive or negative from one month to the next. Task 4, however, showed cumulative curves, which could represent the current value of various equities from month to month, meaning that the slope of these curves was meaningful, and also allowing the curves to stray more and more from the $y = 0$ baseline as more time passed.

Notice that Tasks 1 and 2 required comparison of individual curves, whereas Tasks 3 and 4 required comparing entire groups of curves.

All four tasks have analogues in financial data analysis : Task 1 asks the user to compare three particular equities that have been somehow selected a priori, and are thus highlighted visually.

Task 2 asks the user to identify the equities which performed the best (and are therefore not yet selected nor highlighted) and also estimate how much better the best performed compared to the worst of the three. Task 3 asks which sector is the least volatile, which is important for reducing financial risk. Task 4 asks which sector performed the best over a given period.

Notice that none of our tasks involve intra-group comparisons, nor questions about a single group. This is because such tasks are probably best done by first filtering out or hiding the other groups, to reduce the number of curves and complexity of layout, in which case there is only one group left to display, and there are no interesting differences between the various layout techniques we consider if there is only one group to display.

For each task, participants were asked to perform the same steps for all 6 techniques times 4 trials. The ordering of techniques was counterbalanced with a 6×6 Latin square. In all tasks, at the start of each trial, the user was shown the curves, immediately below which was the question, along with radio buttons and (in the cases of Tasks 2 and 3) a text field. After selecting a radio button and possibly entering a range in the text field, the user clicked on a "Next" button to complete the trial and move on to the next trial.

To increase control over the experimental conditions, we evaluate the charts with no interaction, e.g., no interactive selection or highlighting of curves, no brushing and linking, and no interactive highlighting of specific X or Y values. However, in Task 1, to make it easier to communicate the task to the user, we highlight the curves to be compared. Because these are outlier curves (i.e., curves suffering from almost no occlusion), they would be easy to notice and perceive in a real world scenario without highlighting them, and highlighting them should not make it any easier to compare their shapes or values ; highlighting them in the experiment only serves to identify them for the task.

3.5.4   Results

3.5.4.1   Task 1

In this task, since curves were highlighted and clearly visible, conventional was expected to outperform the others. The task was basically created in order to evaluate compression and shifting costs against the state of the art.

Table 3.1 presents the results for task 1. As expected, conventional presents better results than any other technique. We performed a chi-square test of independence with the success rates, and ANOVA combined with pairwise t-test with Bonferroni corrections when $p < 0.05$ on log transformed times. Success rates and techniques were proven dependent (χ^2=64.19, $p < 0.001$), mostly because of VSM. Time also differ significantly ($F_5 = 21.48$, $p < 0.001$). Superposed was not significantly slower than conventionnal ($p = 0.62$), but juxtaposed ($p = 0.03$) and shifted ($p = 0.01$) were slightly slower.

Small multiples were both significantly slower than conventional ($p < 0.001$) and generated more errors, especially VSM. All participants selected VSM as the worst technique for this task. Since each technique uses a different y axis, participants have to evaluate y values for each layer and compare the results mentally. They also found it more difficult to obtain a precise estimate of the value since VSM is heavily distorted vertically.

5 out of the 12 participants preferred conventional for this task, and 5/12 preferred superposed. This result is especially surprising since noise was not an issue in this task. Still, in some cases, since the highlighted curves were drawn with a large stroke and a white contour, some occlusion could appear when curves were close enough. Using superposed, highlighted curves were never drawn on top of each other and values were easier to compare in such situations.

Tableau 3.1 Task 1 results

Technique	Success rate	Avg. time (s)	SD time (s)
Conventional	99%	14.2	0.47
Superposed	95%	17.0	0.45
Juxtaposed	96%	18.3	0.43
Shifted	94%	18.2	0.37
HSM	92%	21.2	0.33
VSM	70%	33.7	0.51

3.5.4.2 Task 2

Task 2 first required the user to find the most performing stock. This curve was not highlighted but was clearly visible. We expected similar results to the first task for this part. Estimating the range between all three maxima was expected more to be a challenge since the maximum might not be easily detectable for all layers (because of occlusion), and comparing the interval might be a challenge with small multiples, for example, because of the disconnected axes.

Table 3.2 presents results for task 2. Most techniques presented similar success rates to the first task, except VSM. One participant gave us a hint about the reason why that could be the case : "I felt that VSM was less painful since I already had to measure the Y values for the second part of the question." Since participants were forced to measure the values, they did not try to guess the answer simply by looking at the curves. In this case, success rate and techniques are not dependent (χ^2=3.78, $p = 0.73$).

Interval errors, on the other end, differ significantly based on the selected technique (ANOVA on log transformed errors, $F_5 = 15.13$, $p < 0.001$). Juxtaposed appears to be the most precise technique, and offers a significantly smaller estimate error than all other techniques ($p < 0.001$). The noise level is reduced compared to most other techniques since layers are not over-

layed. In certain cases, participants missed the maxima on the inner layers using compressed, shifted and conventional because they were hidden in the background noise, resulting in large estimate errors. Layers are also presented close to each other using juxtaposed, which also helps comparing the values. Time was significantly different between techniques ($F_5 = 2.56$, $p = 0.03$), but only VSM is significantly slower than conventional ($p = 0.0466$) based on the Pairwise test.

Compared to task 1, 8/12 participants selected a different technique as their preferred. Participants were mostly divided between juxtaposed (5/12) and superposed (5/12). People preferred juxtaposed because of the complete layers separation, and superposed because layers are positioned closer to each other. Although the results were much better than for task 1, 10 participants still selected VSM as the worst technique. Conventional and HSM also collected one vote each for the worst technique.

Tableau 3.2   Task 2 results

Technique	Success rate	Avg. error	SD error	Avg. time	SD time
Conventional	100%	5.69	1.29	30.1	0.43
Superposed	98%	5.99	1.37	35.0	0.37
Juxtaposed	96%	2.13	5.60	31.4	0.42
Shifted	94%	4.26	1.02	31.3	0.41
HSM	96%	6.69	1.17	33.2	0.35
VSM	98%	4.59	1.01	36.9	0.36

3.5.4.3   Task 3

The third task was also composed of two subtasks. The first one was to correctly identify the sector with the smallest volatility. Then, the participant was asked to evaluate the interval

between the highest and the lowest values for that sector. This task was specially created to evaluate the impact of occlusion.

Table 3.3 presents the results for the third task. Compressed and juxtaposed clearly underperformed for identifying the correct layer (χ^2=27.32, $p < 0.001$). Some people found juxtaposed especially difficult when the layers were not aligned (similar intervals with different maxima/-minima). Juxtaposed tends to create tall vertical columns when values are spread on the y-axis, and comparing such columns might be challenging because of the resulting distortion on the x-axis.

Error rates varied significantly between techniques ($F_5 = 3.61$, $p = 0.0035$). Superposed presents some surprising results. Participants made more errors in identifying the right sector, but were able to evaluate the intervals significantly more precisely ($p = 0.0011$) than conventional. When looking at individual trials, we realized that all mistakes, except one, were made in the same trial (5/12 failed the trial). Two layers had similar intervals and it would have been very difficult to get the correct answer using any technique. If we removed this trial, success rate would be 97.2%. If we consider this correction, superposed would clearly be the best technique for this task.

Subjective preference for this task is much more divided than for the previous tasks : juxtaposed (4 votes), compressed (3 votes), shifted (3 votes) and VSM (2 votes) shared the preference. Shifted also performed quite well in this experiment, although some participants complained that it was especially difficult when the amount of curves was uneven between layers. Many participants mentioned that VSM was the easiest technique to find the less volatile sector, but was harder to evaluate precisely the range. Juxtaposed was also especially appreciated for the first part of the task.

7 participants disliked conventional since it was often very difficult to answer because of occlusion. In many cases, participants guessed that one layer was the smallest since they barely saw it. VSM (3 votes) and HSM (2 votes) collected the rest of the votes.

Tableau 3.3   Task 3 results

Technique	Success rate	Avg. error	SD error	Avg. time	SD time
Conventional	100%	6.51	0.98	40.6	0.40
Superposed	88%*	2.77	1.72	36.3	0.37
Juxtaposed	81%	4.35	1.01	38.4	0.38
Shifted	100%	4.10	1.12	35.2	0.33
HSM	100%	4.58	1.48	35.7	0.33
VSM	96%	6.57	1.29	33.5	0.50

*One trial presented two layers with almost the same interval. 5/12 missed the trial. Without this trial, the rate would be 97.22%.

3.5.4.4   Task 4

The last task tries to evaluate the impact of compression on slopes evaluation. Since a different level of compression is applied to each layer with superposed, we excepted this technique to underperform. Evaluating slopes on large periods could also be a problem with compressed techniques because of the discontinuity between the ticks.

Table 3.4 presents the results. Against all odds, superposed appears to present the best results, although success rates are independent (χ^2=8.35, $p = 0.14$). 4 participants preferred this technique for this task. Curiously, VSM also collected 4 votes, and presented the worst results for this task. Many participants commented that VSM gave them the best picture to guess the average slopes, but ended up giving invalid answers in most cases. Because of the small space available for each layer vertically, VSM curves suffered greatly of the resulting distortion and were more difficult to interpret. Times were not significantly different ($F_5 = 1.70$, $p = 0.1352$)

Half of the trials were executed on a single month interval, and the other half on multiple months. Table 3.5 shows the success rates for the two cases. Comparing multiple months was clearly more difficult and compression was expected to suffer more of the disconnections and

Tableau 3.4　Task 4 results

Technique	Success rate	Avg. time	SD time
Conventional	58%	20.4	0.45
Superposed	67%	23.1	0.50
Juxtaposed	50%	23.3	0.39
Shifted	44%	24.5	0.45
HSM	52%	22.4	0.48
VSM	42%	19.7	0.44

distortion. Surprisingly, superposed and juxtaposed did better than all other techniques, and shifted, which was introduced to remove those factors, performed the worst.

Tableau 3.5　Task 4 success rates for different intervals

Technique	Single month	Multiple months
Conventional	67%	50%
Superposed	71%	63%
Juxtaposed	46%	54%
Shifted	63%	25%
HSM	63%	42%
VSM	38%	46%

3.5.4.5　Global

We calculated global statistics for all four tasks, as show with table 3.6. Although conventional presented a slightly better success rate than our proposed techniques, our techniques were

generally more precise than the state of the art ($F_5 = 8.31$, $p < 0.001$). Juxtaposed performed significantly better than all other techniques globally ($p < 0.001$). They also collected all the votes for global appreciation, with compressed (6 votes), juxtaposed (4 votes) and shifted (2 votes). In many cases, participants admitted that superposed was not always the best technique for all tasks, but always worked well. It also collected the best scores for ease of use, speed, precision and satisfaction. Participants only preferred conventional for it's learning curve. Table 3.7 presents the global appreciation scores for all techniques.

We compiled user comments for each technique and summarized them in table 2.1 (presented in appendix II).

Tableau 3.6   Global statistics

Technique	Success rate	Avg. error	SD error
Conventional	91.25%	6.10	2.401
Superposed	88.33%	4.38	1.997
Juxtaposed	83.75%	3.24	4.3473
Shifted	85.00%	4.18	1.0652
HSM	86.25%	5.64	1.3536
VSM	75.00%	5.58	1.1497

3.6   Discussion

Based on the results presented in the previous section, we can state that compression can outperform state of the art techniques for certain tasks. Although superposed appears to be the most appreciated technique and in average the most performing one, juxtaposed did score better to compare Y values between layers. Clearly, there is no one size fits all technique. For example, if there is not a lot of curves within each layer, conventional could be the best technique to fit your needs. But in the case where you need to compare a small number of layers

Tableau 3.7　Global Appreciation

Technique	Ease of use	Speed	Learn. curve	Precision	Satisfaction
Conventional	2.75	2.75	4.67	2.42	2.67
Superposed	3.58	3.67	3.67	3.50	3.75
Juxtaposed	3.33	3.42	3.58	3.42	3.25
Shifted	2.75	2.75	3.08	3.17	3.17
HSM	3.00	2.17	4.25	2.50	2.25
VSM	3.08	2.95	3.85	3.00	3.02

containing many hundred of curves, compression could be of interest. Tasks were selected to evaluate use cases where we thought our techniques could be of interest. For example, tasks such as comparing two curves on a single layer would not take advantage of the benefits of compression nor shifting. The added cognitive cost would clearly be a disadvantage compared to the simple conventional. Still, task 1 clearly illustrated that this cost is not significant in most cases.

Superposed layers performed surprisingly well in task 4. Since this technique introduces a different slope ratio for each layer, we would have expected this technique to be the worst. In fact, juxtaposed and shifted were mainly introduced because of this uneven distortion with superposed. Results showed that this technique was the best to evaluate slopes for both single month and multiple months periods, although not significantly better. We can at least confirm that compression doesn't affect slopes analysis.

The shifted technique was much less appreciated in the context of the experiment, and globally presented less promising results. The transparency was considered by many as noise, and the technique was difficult for most participants to understand. Although compression introduces discontinuity and distortion, we think the results look more promising for the stated tasks and would recommend using it instead of shifted.

Although there is no theoretical limit to the number of curves that could be represented with compression, our own experiments showed that 2 or 3 was the ideal, and that more than 5 was very difficult to read. Redundant color coding is also a must, especially with superposed. Even for juxtaposed, color helps connecting the layers together. Choosing clearly distinctive colors that are colorblind safe is highly recommended.

Conventional was rendered with curves sorted by layer. By using this technique, occlusion was clearly a factor of the layer level. One could argue that results might be different if curves were rendered in a more random order. In fact, layers tend to be much more difficult to observe and all the curves tend to blend. Occlusion would still be a problem, and it would not be possible to conclude on any layer. We argue that sorting curves by layers helped the user in this case.

We decided to present VSM and HSM in the same space than all the other techniques. By doing so, we introduced some distortion that would clearly have an impact on the results we observed. Still, we took the stand that a designer has a limited space to affect to a visualization, and that small multiples would necessarly end up with less space than conventional approaches. This is an expected limitation in the real world. Also, we did not want to introduce scrolling as it would have introduced some bias when evaluating answer time.

3.7   Conclusion

We have presented three novel visualization techniques for time series data : two kinds of compressed layers (superposed and juxtaposed), and shifted layers. We also presented the design rationale behind these techniques, and reported an experimental comparison with three status quo techniques, using tasks that have a basis in real-world financial data analysis. In each of Tasks 2 and 3, one of our techniques was best (as judged by average error), and in Task 4, superposed was best (as judged by success rate). Over all tasks, our three techniques resulted in the lowest average errors. We were also surprised that the superposed technique yielded the highest success rate in Task 4, a task requiring users to estimate slopes, despite the fact that the superposed technique distorts slopes with a different x/y scale for each group of curves.

Although there is no single "winner" across all tasks, we have established that the status quo techniques are sometimes not best, and that further investigation into these or hybrid techniques is warranted.

3.8 Future Work

Many of the display parameters in our techniques could be tested to find their optimal values. For example, we would like to investigate the effect of different curve thicknesses, and perhaps develop a rule of thumb or algorithm for automatically choosing curve thickness that increases visibility without aggravating occlusion. We could also investigate the effect of aspect ratio of charts on different tasks. For example, one rule of thumb (Heer et Agrawala, 2006) states that line graphs have a scale chosen so that the average slope of curves is 45 degrees. We suspect that changing the aspect ratio of charts could have an effect on tasks like Tasks 3 and 4. Shifted layers might also be optimized with different opacity levels or gradient patterns, or with a different spacing between layers. Finally, we would like to evaluate interactive techniques, like brushing and linking, animated transitions between the different techniques, and filtering.

CHAPITRE 4

**VARIATIONS ON TERNARY PLOTS FOR VISUALIZING PORTFOLIO
PERFORMANCE ATTRIBUTION**

Maxime DUMAS[1,2], Michael J. McGUFFIN[1], Steeve Pratte[2], Patrick CHASSÉ[2]

[1] Département de Génie Mécanique, École de Technologie Supérieure,

1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

[2] Croesus Finansoft, 600 Armand-Frappier, Laval, Québec, Canada H7V 4B4

4.1 Abstract

Ternary plots illustrate the proportion of three variables that sum to a constant, usually 1 (or 100%). We designed and evaluated two novel visualisations based on the classic ternary plot in the context of financial portfolio performances analysis. Our financial model decomposes performance differences against a benchmark into three different explanatory effects : stock selection effect, asset allocation effect and currency effect. Our plots allow us to visualize those effects against the excess return and to pinpoint precisely what happened and, more importantly, why it happened, on a large subset of portfolios. We developed a fully functional analytic dashboard featuring our novel charts based on ternary plots. We evaluated our prototype with finance professionals using a real dataset of portfolios. Using our tool, experts were able to detect unexpected anomalies and even reflect on the strategies used by individual financial advisors.

4.2 Introduction

Losing 5% of an investment is rarely considered an happy outcome. But what if you were in the middle of a financial crisis and the global market had lost 25% during the same period ? You might judge yourself to be skilled (or lucky) by having succeeded in limiting your losses !

Investment performances should always be considered against a benchmark. For example, you could use the S&P 500 Global Index as a reference point. The delta between your benchmark and your own investment return is called differential return (or excess return).

If you have a diversified portfolio containing many stocks, bonds and foreign investments, understanding what caused the differential return might prove to be a challenge. *Performance attribution* tries to provide a structured answer to this question using mathematical models. Applying these models on large datasets involves considerable efforts and is error prone.

Following an extensive survey of the finance visualization literature (Dumas *et al.*, 2014), we did not find any contribution addressing the use of visualization to aid portfolio performance attribution. Many works measured performances and allowed for easy comparison between securities or portfolios, but none tried to explain portfolio performances deviations using attribution. The current work is the first to address this gap in the literature.

In our work, we designed two novel charts based on the ternary plot to combine in a single chart the attribution effects and the differential return. We also developed an attribution analysis dashboard around these novel charts. Finally, we evaluated our solution on a real dataset with four experts from the field.

4.3   Related Work

Information visualization is widely used in finance. For example, technical analysts use line graphs, candle stick charts and other relatively simple charts to represent mostly historical time series. Still, there is limited literature about advanced visualization systems and techniques in finance. During the "Challenges in Financial Visualization" panel at the IEEE 2014 VIS Conference in Paris, experts from worldwide financial institutions pleaded for more advances in the field to help them cope with extremely large and complex datasets, while admitting that users were reluctant to use new and unproven techniques.
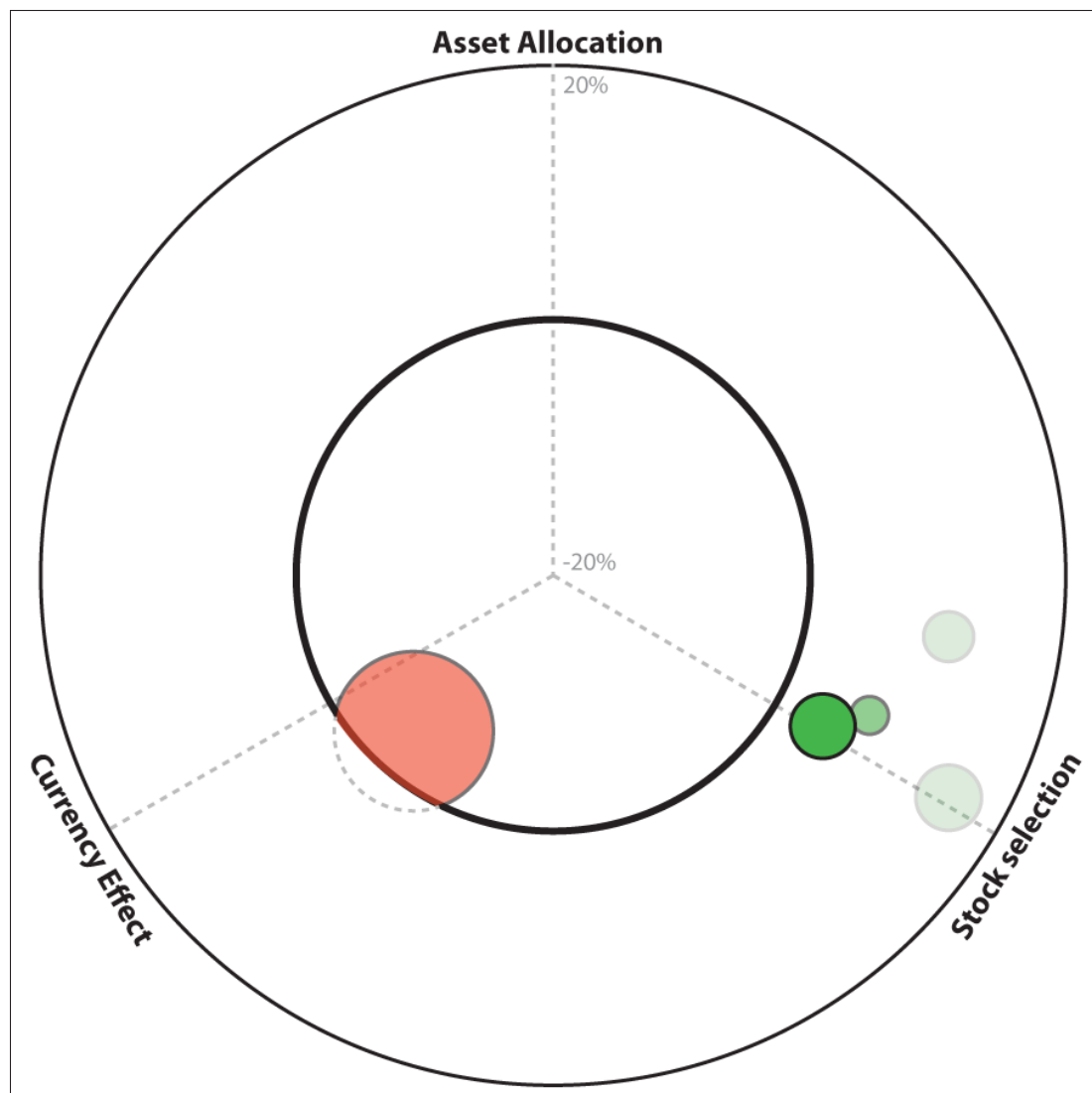
Figure 4.1   The ternary wheel, one of our ternary chart variations proposed
for portfolio performance attribution.

We published at the same conference an initial survey of academic contributions in the field
(Dumas *et al.*, 2014). Among those, few techniques concerned performance analysis. Ziegler
et al. (2008) proposed a pixel visualization allowing the comparison of historical returns for
a single stock for multiple years based on more than 4500 buy/sell date combinations. They
also proposed (Ziegler *et al.*, 2010) a pixel bar chart variation and a spark line matrix to com-
pare individual stocks in different industrial sectors or countries. Alsakran *et al.* (2009) used
heat maps to compare mutual fund returns based on different characteristics, such as fees,

fund size, etc. Marghescu (2007) proposed different techniques, including parallel coordinates, scatter-plot matrices and star glyphs, to compare stocks fundamental values and their returns. Self-Organizing Maps (SOM) are also often used (Eklund *et al.*, 2003; Lansiluoto *et al.*, 2004; Sarlin et Eklund, 2011) to evaluate and compare different industries using financial benchmarking. All of these techniques are limited to individual securities, and do not address portfolio performance analysis.

Savikhin *et al.* (2009) proposed a visualization system to help investors compare returns and risks for different portfolio alternatives while creating their own. Dwyer et Gallagher (2004) also proposed advanced 3D techniques to compare multiple portfolios and evaluate portfolio changes over a certain period. Although interesting, these techniques would not scale to the amount of data used in real life, and do not address performance attribution.

Ternary plots, also known as triangle plots, De Finetti (1926) diagrams (in population genetics) and simplex plots (in game theory), display the proportion of three variables that sum to a constant on a triangular chart. These charts are mostly used in physical sciences to represent relations between three elements composing a system. Weisbuch *et al.* (2000) used a ternary plot in their finance related paper to illustrate trading behaviors between three agents on perishable goods markets. D'Urso *et al.* (2013) also used this chart to represent relations between clusters of currency rates with the Euro currency. We are not aware of any work using ternary plots to analyse portfolio performances.

4.4 Performance Analysis

Performance analysis can be decomposed in three majors stages (Lawton et Jankowski, 2009) : performance measurement, performance attribution and performance appraisal. Performance measurement is the most common and obvious phase of all three. It involves calculating the portfolio return for a specific period, and to evaluate how each position contributed to the overall portfolio. Most finance visualization related publications in our survey (Dumas *et al.*, 2014) involve this stage.

Performance attribution tries to explain differential returns against a benchmark (i.e. the market) and their impact on the portfolio. A simple return by itself doesn't give much information. You might have lost 5% during a month, but if the market suffered a major drop of 15%, your performance was not that bad after all! Performance attribution analyses that 10% differential and tries to break it into different factors to give insight on what explains the results. For example, is it because the stocks you chose performed better than the ones in the benchmark (stock selection)? Or is it because you invested more in long term bonds than the benchmark and it reduced your exposure to risk (asset allocation)? Or maybe because you invested in US stocks and the US currency worked in your favor (currency effect)? There are multiple attribution models available depending on the level of granularity desired and the type of assets under consideration (Bacon, 2008). We chose the Geometric Multi-Currency Attribution (Karnosky et Singer, 1994) model as it provided the level of detail expected from our partner. The model could be formalized by equation 4.1. The three factors on the left-hand side correspond to stock selection effect, asset allocation effect and currency effect, respectively, and when multiplied, result in the total differential return.

$$\left(\frac{1+r_L}{1+b_{SL}}\right) \times \left(\frac{1+b_{SL}}{1+b_L}\right) \times \left(\frac{(1+r)(1+b_L)}{(1+r_L)(1+b)}\right) = \frac{1+r}{1+b} \tag{4.1}$$

The parameter r represents the portfolio return, as defined by equation 4.2, with w_i representing the weight of element i in the portfolio, and r_i the return of this element.

$$r = \sum_{i=1}^{1=n} w_i r_i \tag{4.2}$$

The parameter r_L represents the portfolio return converted in local currency. For example, if your investment in CAD rose 2%, but during the same period, your local currency (USD) dropped by 5% against CAD, your local return is -3%. Equation 4.3 defines this concept, with r_{Li} being the local return of the ith element.

$$r_L = \sum_{i=1}^{1=n} w_i r_{Li} \tag{4.3}$$

Similarly, you get the benchmark return b and local return b_L using similar formulas, but weighting by W_i, the weight of the ith element in the benchmark. Finally, b_{SL} represents the benchmark semi-notional fund, which represents the sum of benchmark local returns times the weight of the element in the portfolio (instead of the benchmark). Formula 4.4 defines this concept.

$$b_{SL} = \sum_{i=1}^{1=n} w_i b_{Li} \tag{4.4}$$

We realized that taking a logarithm of both sides of equation 4.1 yields a model amenable to a ternary plot, as defined by equation 4.5.

$$\log_{\frac{1+r}{1+b}}\left(\frac{1+r_L}{1+b_{SL}}\right) + \log_{\frac{1+r}{1+b}}\left(\frac{1+b_{SL}}{1+b_L}\right) + \log_{\frac{1+r}{1+b}}\left(\frac{(1+r)(1+b_L)}{(1+r_L)(1+b)}\right) = 1 \tag{4.5}$$

Based on this formula, 100% of the differential return can be expressed as the sum of the log transformed stock selection, asset allocation and currency effects.

The last stage of performance analysis is performance appraisal, which tries to explain whether the observed differential returns are the result of skill or luck. Although interesting, this concept is out of scope for this project.

4.5   Ternary Plot

Since our model can be represented as three variables that sum to a constant, the ternary plot is an interesting way to represent the information. It allows encoding the three parameters in a simple two-dimensional triangle.

Let's define a as the asset allocation effect, b the stock selection effect and c the currency effect, where $a + b + c = 1$.

The three variables can be projected in a ternary plot with a unit edge triangle using formula 4.6. It was chosen to map $(a, b, c) = (1/3, 1/3, 1/3)$ to the origin $(x, y) = (0, 0)$, coinciding with the triangle's centroid.

$$[x, y] = \left[\frac{a + 2b}{2} - \frac{1}{2}, \frac{\sqrt{3a}}{2} - \frac{1}{2\sqrt{3}} \right] \tag{4.6}$$

A ternary plot can also be read by calculating segment ratios, as shown by the three equations 4.7. Figure 4.2 defines the segments used in the equations 4.7.

$$a = \frac{MX}{MA} \qquad b = \frac{NX}{NB} \qquad c = \frac{LX}{LC} \tag{4.7}$$

The ternary triangle is appropriate to display the relation between the individual effects. In the case where point X would be located near point A, it could be interpreted as meaning the excess return is mainly explained by the asset allocation effect. In the case where the point would be located at the centroid, it would mean that all effects have an equal weight in explaining the differential return.

4.6 Ternary Plot Variations

Before trying to understand why a differential return was observed, the analyst must figure out which portfolio presented abnormal returns. The ternary plot, in this case, only tells part of the story. The analyst might not care to know what factor accounts for a 0.1% percent difference, for example. With this idea in mind, we developed variations on the ternary plot in order to display simultaneously the differential return and the dominant effects on a single chart to identify patterns and anomalies in large datasets.
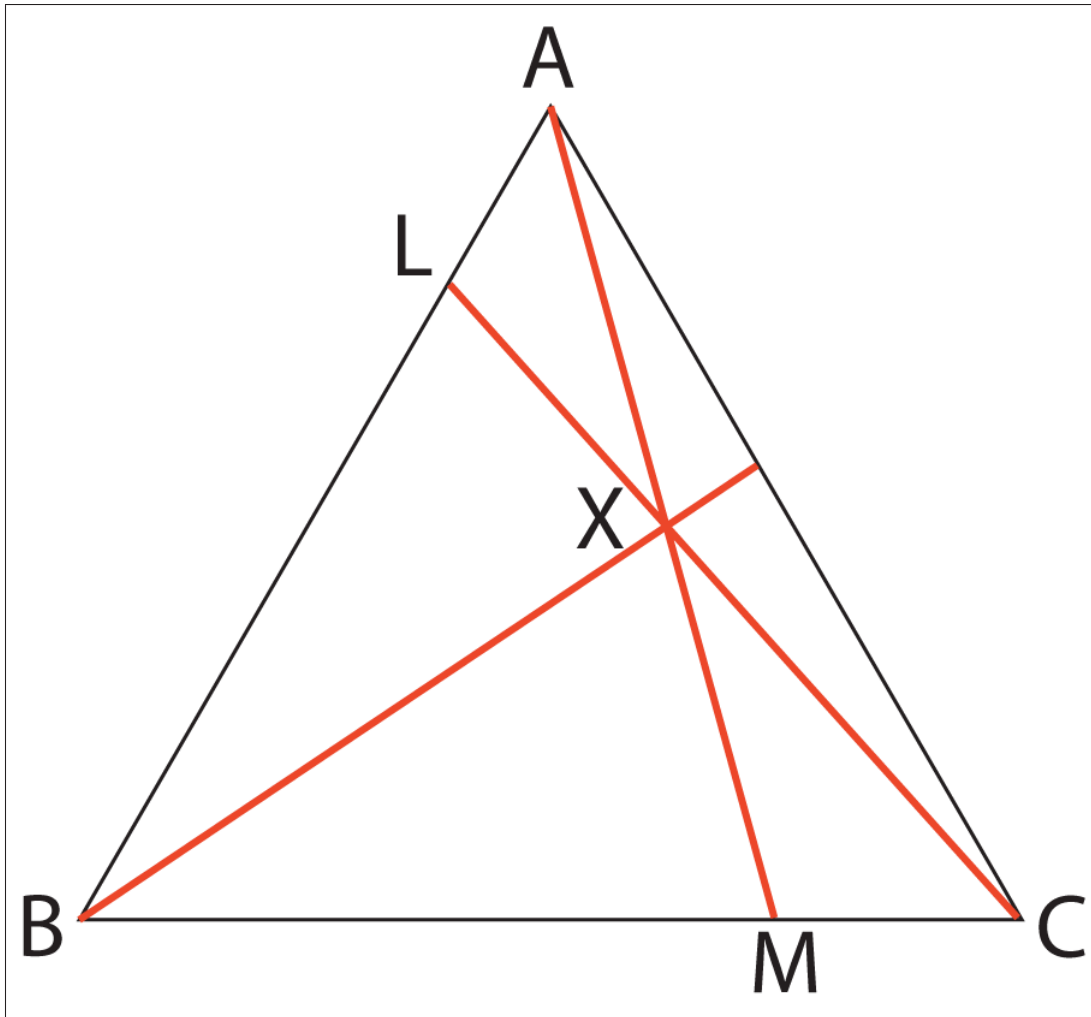
Figure 4.2   Variable proportions can be obtained by calculating segment ratios inside a ternary plot.

4.6.1   Ternary Navigation Bar

The ternary navigation bar (figure 4.5) gives an overview of a large amount of portfolios and simplifies navigation through the dataset.

Although the navigation bar concept is strongly related to the ternary plot, no triangle is displayed by default. Instead of displaying the point in the triangle, a color is derived based on the position in the ternary plot and is displayed as a colored cell to the user.

Many color spaces are based on three primary colors, such as Red/Green/Blue (RGB) and Cyan/Magenta/Yellow (CMY). By filling the ternary triangle using a color space, we end up with something related to a triangle color picker. Using formula 4.6, it is possible to pick a color located at the exact position where the point would end up on the triangle as the position summary.
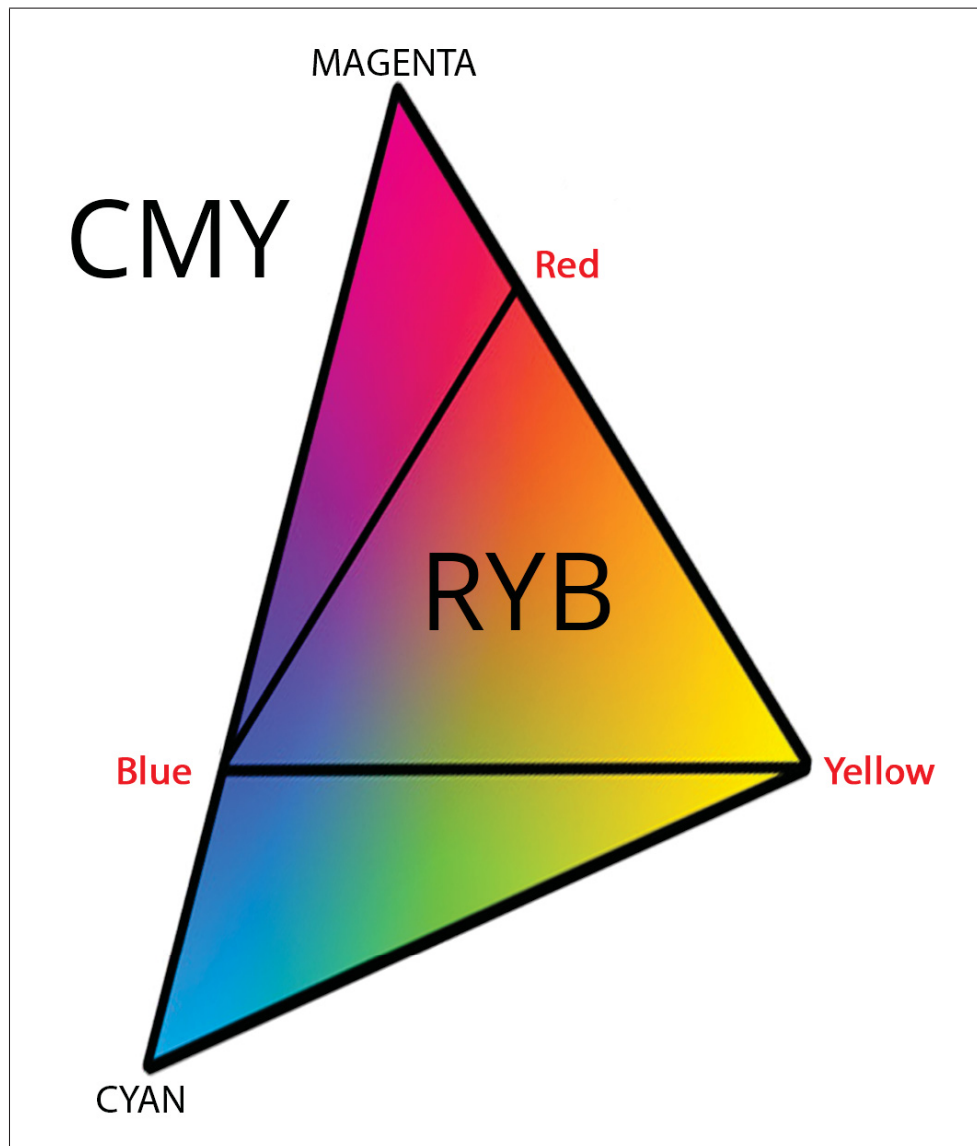
Figure 4.3   RYB is a subset of the larger CMY colorspace.

Mixing mentally RGB or CMY is far from obvious, though. For example, in RGB, red and green would add up to yellow, which is not something we might expect our users to know intuitively. Instead, we took a subset from the CMY color space (figure 4.3) : Red/Yellow/Blue (RYB). This space is easier to imagine mixing as it matches painting behaviors. In this case, yellow and blue would mix as green, and red + yellow as orange. Figure 4.4 shows the resulting triangle with the RYB color space.
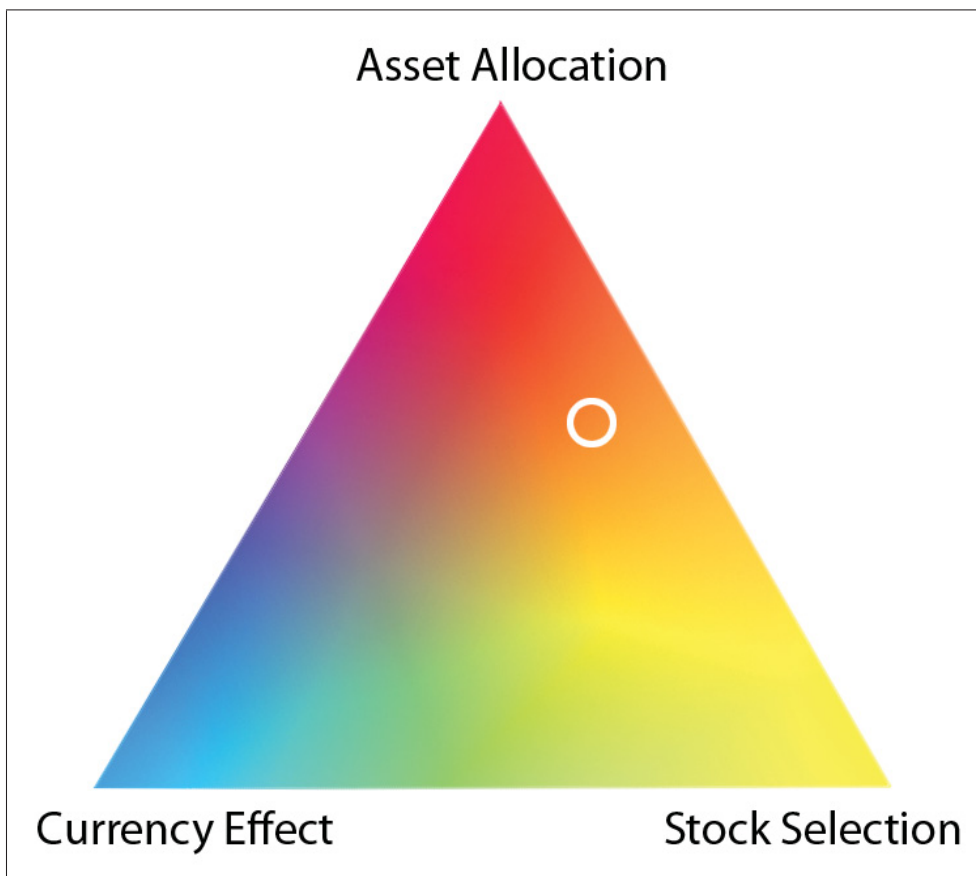
Figure 4.4   RYB ternary triangle. In this case, excess return is mostly explained by asset allocation and stock selection. The Currency effect is weak.

By associating red to asset allocation, yellow to stock selection and blue to currency effect, it is possible to guess mentally which color mix could create the resulting colors, although colors

appear lighter than their painting counterparts (for example, the perfect mix of all three yields a beige instead of dark brown like in painting).

Figure 4.5 illustrates the navigation bar concept. It contains two rows and multiple columns. Each column represent a single element in the portfolio. The column width represents the element's relative weight in the portfolio. The first row encodes the differential return using a red-white-green scale. Strong negative values would get a strong red color, near zero values would be close to white, while strong positive values would be strong green. The second row contains the RYB ternary triangle resulting color (the attribution effect mix). As we show later, the user can click on cells of this bar to drill down into a dataset.
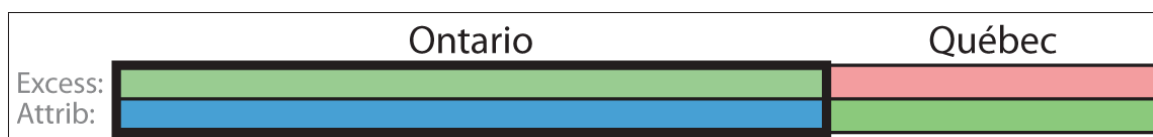
Figure 4.5   Navigation bar concept. In this example, Ontario represents 2/3 of the portfolio, has a positive excess return which is mostly explained by currency effect (blue).

Although this technique might not appear obvious at first, it was well received by the experts. It provides an overview of a portfolio showing what is going on and why. Abnormal situations also tend to stand out with strongly different colors than the rest. Even when not reading the colors to interpret the effect, having one cell with a different color than the rest draws attention to the problematic element.

To help the analysts, a tooltip appears on mouse over displaying the numeric values and the RYB ternary triangle containing the highlighted point (figure 4.6). This technique is also useful when combined with sorting and filtering techniques.

4.6.2   Ternary Wheel

The ternary wheel uses a completely different approach to display the differential return. Instead of relaying on color to encode the attribution variables, we perform a polar projection
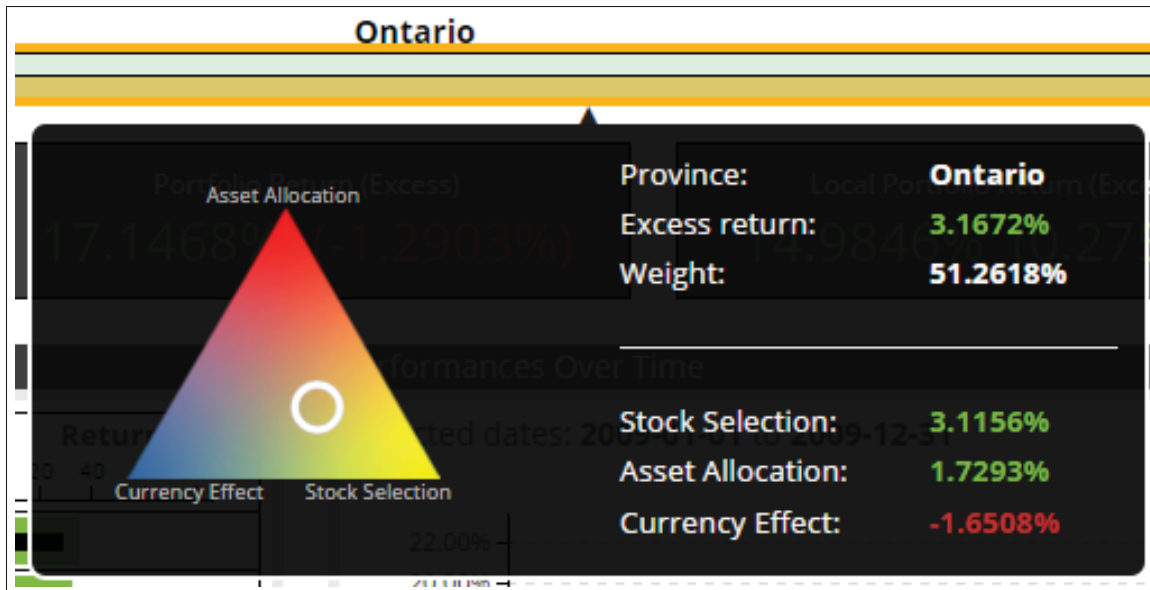
Figure 4.6   A tooltip is displayed with additional information (including the RYB triangle) when the user performs a mouse over on the element cell.

around the triangle centroid to reduce the space to a single dimension. Figure 4.7 illustrates the projection concept.

Formula 4.8 can be used to calculate the projected angle. This angle gives insight into the dominant effects. As shown by figure 4.7, it can then be plotted on a single dimension chart. In this example, a mixture of asset allocation and stock selection explains most of the differential return.

$$\theta = \tan^{-1}\frac{\sqrt{3}(-2a+b+c)}{3(b-c)} \qquad \theta \in]-\pi, \pi[\tag{4.8}$$

Any point on the arrow on figure 4.7 would map to the same projected angle. To account for the missing dimension, we encode the distance from the centroid using opacity (formula 4.9). Using this technique, an opaque circle would mean that the resulting effects mainly (or completely) explain the excess return. On the contrary, a translucent circle would imply that the other effect(s) should not be neglected.
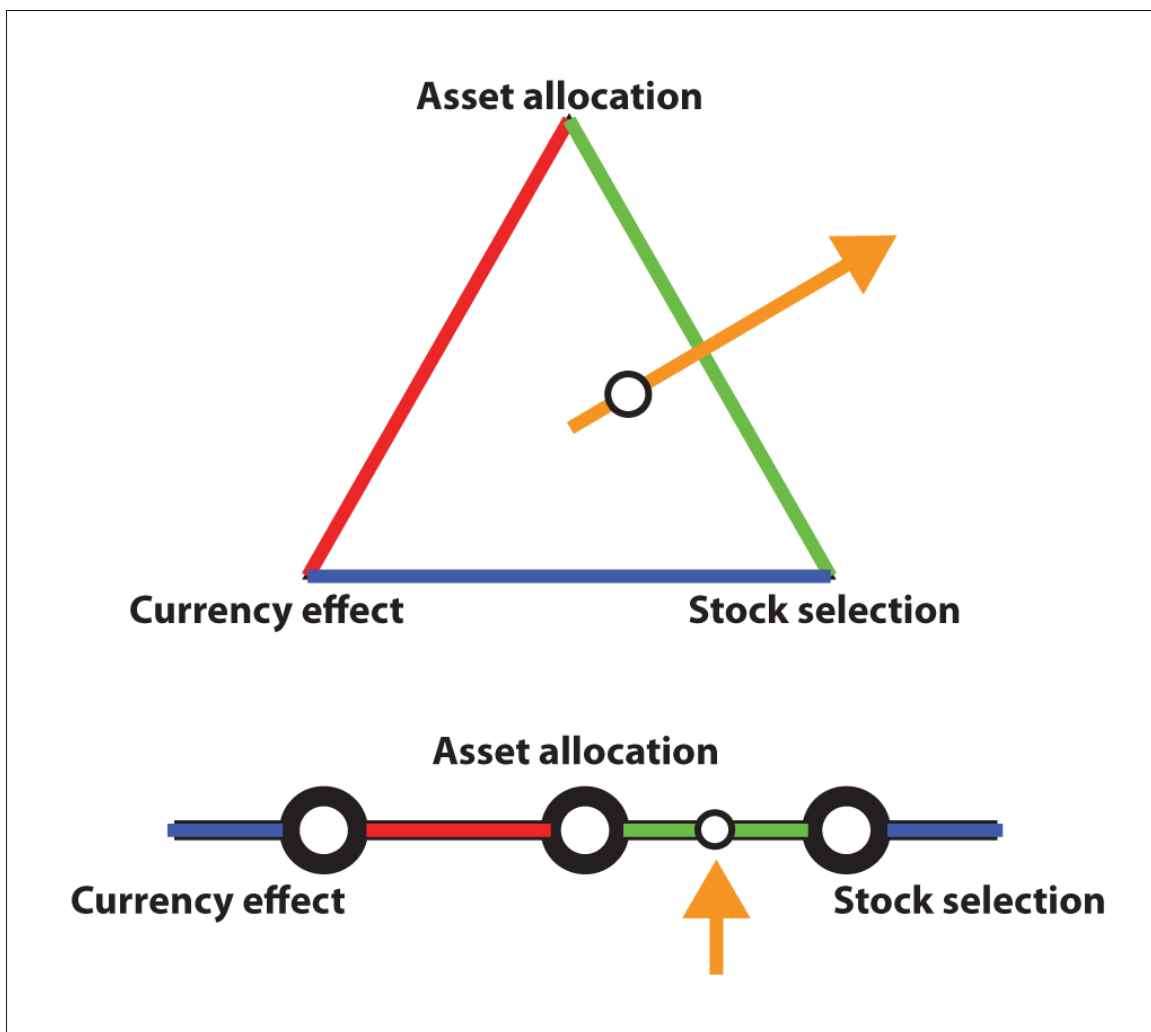
Figure 4.7   Dominating effects can be determined using a polar projection from the triangle's centroid to the triangle's edges. The missing parameter (distance from centroid) will be encoded using opacity.

$$\alpha = \sqrt{a^2 + b^2 + c^2 - ab - bc - ac} \; , \; 0 \le \alpha \le 1 \qquad (4.9)$$

Figure 4.7 shows the projected angle on a simple line. Because of the resulting discontinuity, this technique might not be intuitive to use. Instead, we plot the angle on a circle, as shown by figure 4.8. Effects are placed around the circle similarly to the ternary plot. The radius is then used to encode the differential return, in order to get the full picture on a single chart.

Figure 4.1 gives a complete example with five elements in the portfolio. Circle size represents their relative weight in the portfolio, and are color-coded red or green based on their excess return. In this example, most elements produced a positive differential return (outperformed the benchmark), which can be explained in most cases with stock selection (better performing stocks than the one selected in the benchmark). Only one portfolio produced a negative excess return, caused mainly by the currency effect (probably invested in a foreign currency that lost value during that period).
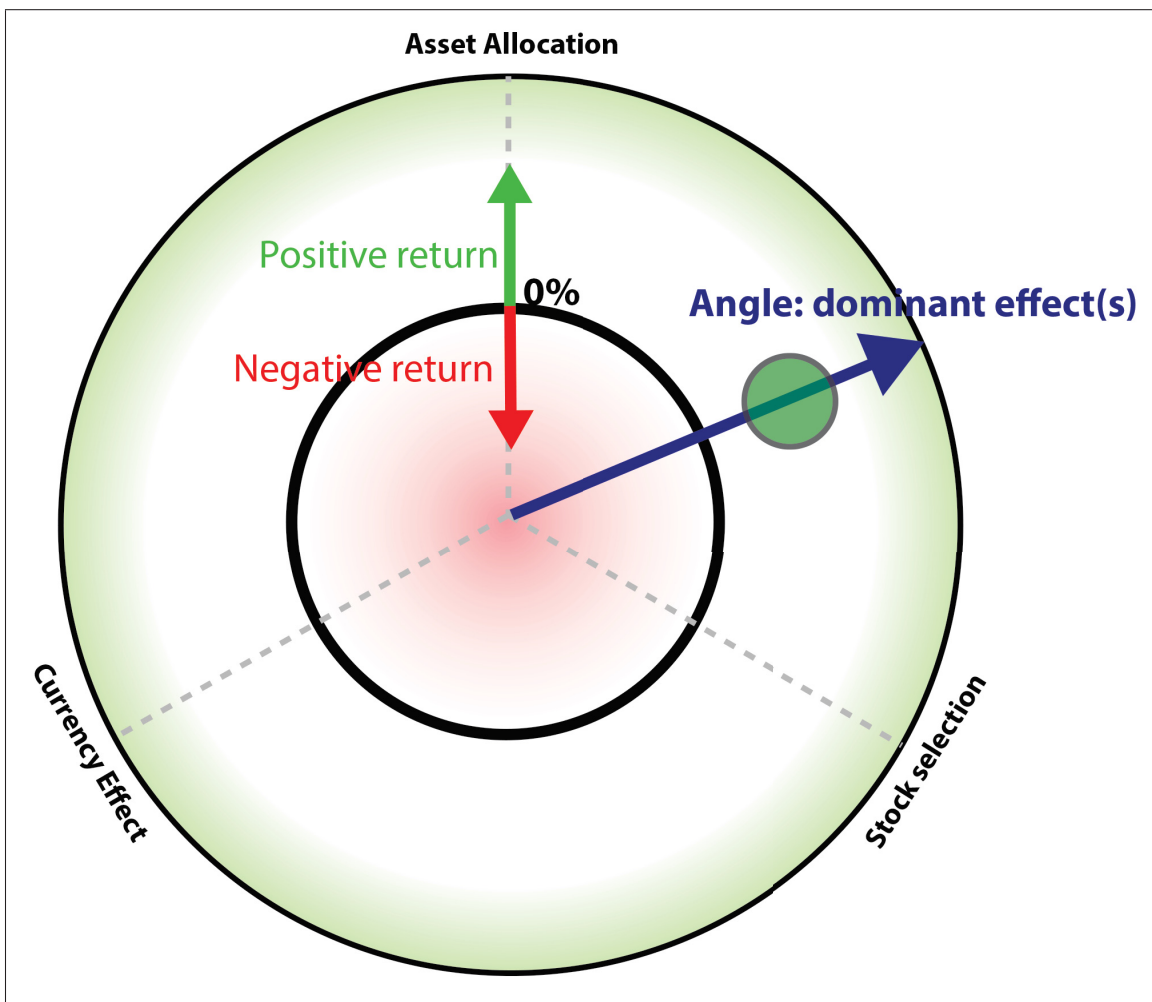
Figure 4.8   Portfolio elements are positioned according to their differential return (radius from inner circle) and dominant effects (angle).

4.7 Attribution Analysis Dashboard

In order to put those concepts in context, we built a complete portfolio performance attribution dashboard. The dashboard was designed iteratively in collaboration with a domain expert with extensive work experience in the field. Our expert had been involved at each step of the design and gave us valuable feedback on how to improve our solution.

The system enables performance analysis for a complete firm, and supports drill down by regions, states or provinces, offices, advisors, clients, down to portfolio level. Our proposed novel charts were complemented by more common data tables, bar charts, bubble charts and line graphs to give all the information required by an expert analyst. Figure 4.9 shows the proposed interface.

Figure 4.9   Proposed dashboard featuring the ternary navigation bar on top and the ternary wheel on the right.

By default, the interface starts at the region level, and displays information for the associated provinces. In order to navigate through the different levels and drill down to the portfolio level, we display at the top two ternary navigation bars. The first one (figure 4.9-A) shows the selected portfolio (selected region) compared to the other sister portfolios (all regions - in this case, we only had one). The second bar (B) shows the child level (in this case, all provinces from the selection region). Both bars give an overview of the current and lower levels and allow the analyst to drill down simply by clicking on the desired cell. For example, if the analyst clicked on a province cell, the interface would be filtered to display only the information about the selected province. The top navigation bar (A) would then present the selected province against all provinces in this region, and the lower bar (B) would display all offices associated with the selected province.

The interface also presents basic overall statistics for the selected portfolio (C). On the left, the portfolio is decomposed by asset categories (D), currencies (E) and industry sectors (F). Relative weights and returns for each categories (large gray bands) are presented against benchmark (small black bands) using bullet graphs (Few, 2006). Using this technique, analysts can easily locate differences between the expected benchmark values and the resulting portfolios.

The center section of the dashboard (G) displays on a line graph the evolution of the selected portfolio performances during the selected time period (orange line) against the benchmark (blue line). It also shows the child spread (min/max monthly performance values of all child elements - in this case all provinces) in light red when the spread increased during the month, or light green otherwise. Analysts can also interact with the chart to select a different sub-period. Performances are presented monthly, and daily when a single month is selected.

Underneath, table (H) details all child elements (in this case, provinces) in a similar fashion to the other tables presented previously, with the addition of the attribution column as simple bar charts. The bars match the RBY ternary triangle colors. The same colors are also repeated in the attribution cells in (C) to help the analysts associate the colors with the effects.

The "performance impact by position" table (I) lists all positions (all securities) contained in the selected portfolio. It is mainly used when an anomaly is observed with a dominating stock selection effect. As requested by our partner, values are presented numerically and support filtering in a more advanced way (i.e. individual column search) compared to other tables.

Excess returns and weights (against the benchmark) are plotted in the bubble chart on the right (J). It basically presents the same information as in (D), (E) and (F) in a different way. Instead of presenting the absolute values for both categories and benchmarks, it displays the relative differences in weights against returns. Similarly to the ternary wheel, circle size represents the relative weight in the portfolio, and color highlights a positive/negative contribution.

Finally, the ternary wheel (K) completes the dashboard by displaying all child portfolios (in this case, the two provinces).

All views are coordinated with synchronized highlighting on mouse over. Tooltips also appear giving additional information on demand. The dashboard supports filtering by clicking on an element in any view.

4.8 Expert review

Since the project was sponsored by an industrial partner, we had access to a real dataset and experts from the field. We organized individual interviews with experts and asked them to analyse real portfolios using our proposed dashboard. The purpose of this review was to study the suitability of the proposed techniques in the context of portfolio performance attribution. We used an open-ended think-aloud protocol (North, 2006; Hoeber, 2009) where users were invited to explore the interface and find problems by themselves. Since all of the information was available both in a tabular form and in the charts, we were interested in understanding which tool in the dashboard experts would use to detect and explain anomalies.

4.8.1   Participants

We interviewed four finance experts with different backgrounds. All of them graduated in finance and have a Certified Financial Analysts (CFA) certification or have completed at least the second of three exams of the CFA program. CFA curriculum covers all the theorical aspects of performance attribution required for using our prototype. Two of the experts, including one of the co-authors, also have extensive work experience in performance attribution in different financial institutions. Three of them had never used our prototype before the study and had not been involved in the design process. Only one expert was considered as a technical consultant during the design of our dashboard and had already used the prototype before the formal evaluation. We decided to include him in the evaluation process to evaluate the impact of having an experienced user (would he be able to outperform the other participants because he had a better understanding of the controls ?).

4.8.2   Data Set

We tested our solution using a subset of a real dataset containing portfolios data from a major Canadian bank. Our subset contained 512 randomly chosen portfolios from clients in 4 offices in Canada during five years. Over 220 000 transactions were registered during that period on more than 15 000 securities. Although the dataset may appear to be small, the database still contained multiple gigabytes of data, since performances were calculated daily for each position contained in each account and considering all the daily information required to calculate performance attribution.

Since the data is confidential, an exhaustive process of anonymization was performed on the dataset before we got access to it. Confidential transformations were also applied to the data so that simple correlations could not be performed. Names, dates, locations and benchmarks presented in this paper may not be accyrate, although the basic structure and content of the accounts were preserved to ensure that the analysis still made sense.

4.8.3 Design

We conducted the study with 3 hour individual interview sessions. They were conducted over a period of three weeks and were performed at our partner's offices. We recorded the entire sessions using a microphone and a screen capture tool for later analysis.

Participants were first asked to fill a background questionnaire. Afterwards, the first hour was generally spent reviewing the attribution model used and explaining the proposed techniques and dashboard features. A short quiz was performed after the training phase to ensure that everybody knew where to look for specific information before starting the experiment.

Then, we asked our experts to analyse our dataset at the firm level, like a firm manager would do, for around 30 minutes. We asked them to discuss out loud their observations so that we could follow with them. We did not give any specific guidelines as for where to look in the dataset. We simply pointed out the difference in performance between Quebec and Ontario, and asked them to explain what happened. They were also clearly instructed to dig into anomalies once they found one to understand what caused this performance deviation. We guided them only when we found they were deviating from the expected task.

Then, they were placed at the advisor level and asked to perform the same task for around 30 minutes. This level was technically more obvious for them as it was more related to their own experience. We repeated this condition on two different accounts that presented interesting characteristics.

We also had prepared specific questions to answer in case some of them ran out of ideas while exploring. We asked the remaining questions toward the end of the session to make sure all topics had been covered. Finally, a post-session questionnaire was given to the experts to collect their feedback.

4.8.4 Results

Results greatly differed between participants based on their work experience performing attribution. The two experts who used performance attribution almost daily were very proficient with our tool and knew exactly what to look for, even though one of them had never seen our solution before. They used all the tools at their disposal and were able to discover problems and give plausible explanations. They mastered the tools quickly. Experts with only a theoretical background found the exercise more challenging since they were not exactly sure what to look for to get a deeper understanding of the observed effects. We did provide more help to them and they performed better in the second exercise once they had a better feel of what we were looking for.

For the first task at the firm level, all experts started their analysis using the navigation bar. Since the bar gives an overview of the data, they were easily able to conclude that Ontario outperformed Quebec in our dataset. All effects had similar weights, although with different signs. After browsing through the interface for the whole region, they usually drilled down to a specific province and looked at individual effects, asset classes, sectors and securities.

Since there were at most 2 provinces and 4 offices, most of the remaining analysis was performed with the tables and the bubble chart. The ternary wheel was less of interest with such a small dataset since the table already gave an adequate overview. The navigation bar was mostly the starting point and used to navigate afterwards. Charts were often used to compare elements between the two provinces and to guide the initial intuition. Still, using the interface, experts were able to explain the differences quite easily. Most detected that Quebec overweighted bonds that returned lower returns than the benchmark. Experienced participants detected more subtle elements, such as individual securities that had an important impact on the provinces, or facts like Quebec underweighted the healthcare sector that seriously outperformed the benchmark to explain the stock selection effect. Our collaborator expert was the only one to filter the time period and was able to specifically pinpoint when the problems started.

The ternary wheel was more useful in the second part, at the advisor level, when users were asked to comment on one advisor's client's performance. Since advisors usually had many clients, it became more difficult to conclude on a single client using simply the data tables, even with sorting features. Using the wheel, they were able to conclude directly on the advisor's strategies and management skills simply by looking at the chart. Figure 4.10 illustrates the first specific case given to them. One expert said : "This advisor clearly has a client that makes his own calls, since he clearly stands out from the rest !". By drilling down on this client, he confirmed his thoughts. "He invested all his money in three huge stocks. That's not what an advisor would do !". Another expert commented on the fact that most clients were driven by asset allocation, as shown by the ternary wheel. "The advisor invested using a defensive strategy. He invested more in Canadian stocks and mid-term bonds than the benchmark. It clearly paid off ! Most of the negative results come from his underweighted investments in USD.".

We then switched to another advisor with more clients and a less obvious strategy, as shown with figure 4.11. When asked to comment on this advisor's strategy, all of them looked at the ternary wheel and concluded that he was a passive investor. "This advisor is 'beta' ! He did not create any value !", said one of the experts. "Since all sector weights are almost perfectly equal to the benchmark, and based on the fact that the clients are spread all over the circle, I would conclude that he is not doing any active management, basically following the benchmark !", stated another. Only one client seriously outperformed the others. One expert detected this problem, drilled down and explained that "He only holds 7 stocks, and 75% of his portfolio is invested in one company ! That is usually the case when the client's employer has a stock ownership plan. That's probably why this client is so different from the rest.".

4.8.4.1 Subjective Preference

All experts greatly appreciated their experience with the tool (table 3.7. They confirmed that our solution would greatly enhance current tools. Some experts suggested that the tool would
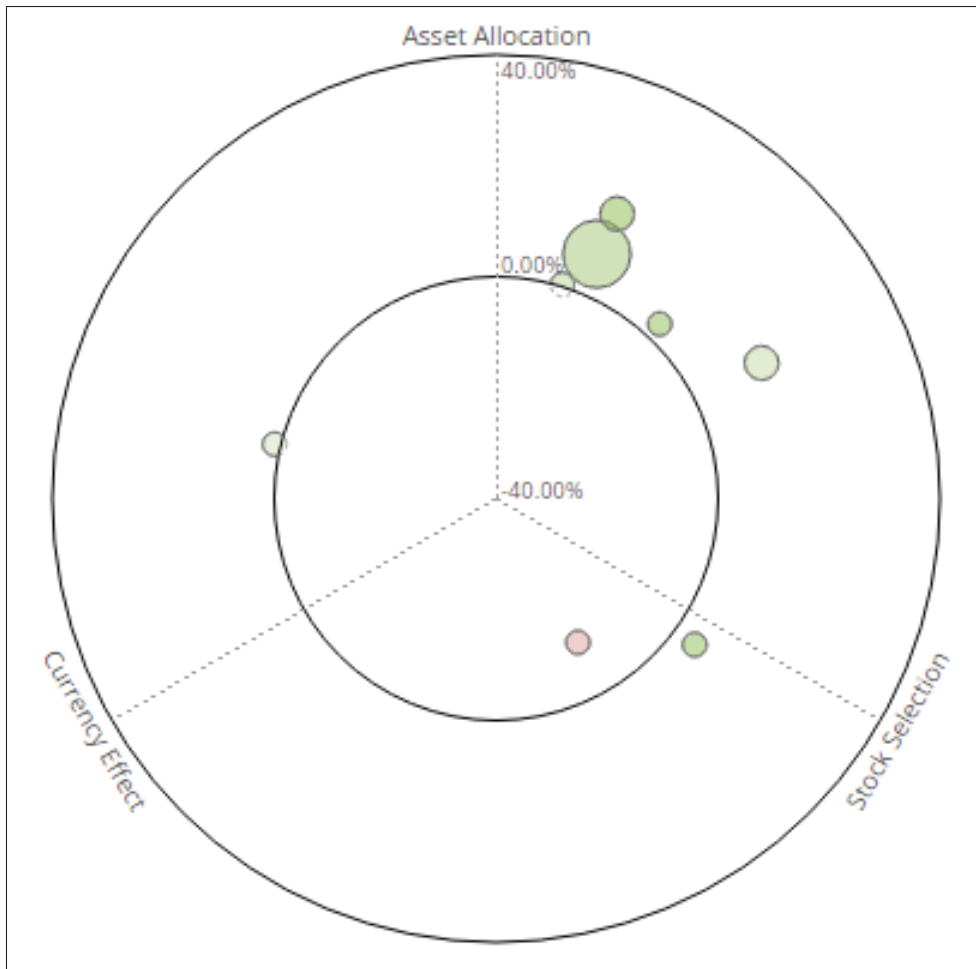
Figure 4.10   The ternary wheel shows a clear strategy where the advisor
outperformed the benchmark by weighting differently the asset classes.

also fit the needs of other professionals (for example, conformity analysts, risk managers, per-
formance teams, etc.).

All experts especially appreciated the ternary wheel at the advisor level. One expert said : "The
circle is less useful at the firm level since there is not enough data. But at the advisor level,
we can clearly see the overall tendencies and even deduce the advisor's strategy !". Another
concluded : "The chart gives a nice overview and brings me to the right effect. Then I simply
have to dig based on what the chart told me. It's really easy !".
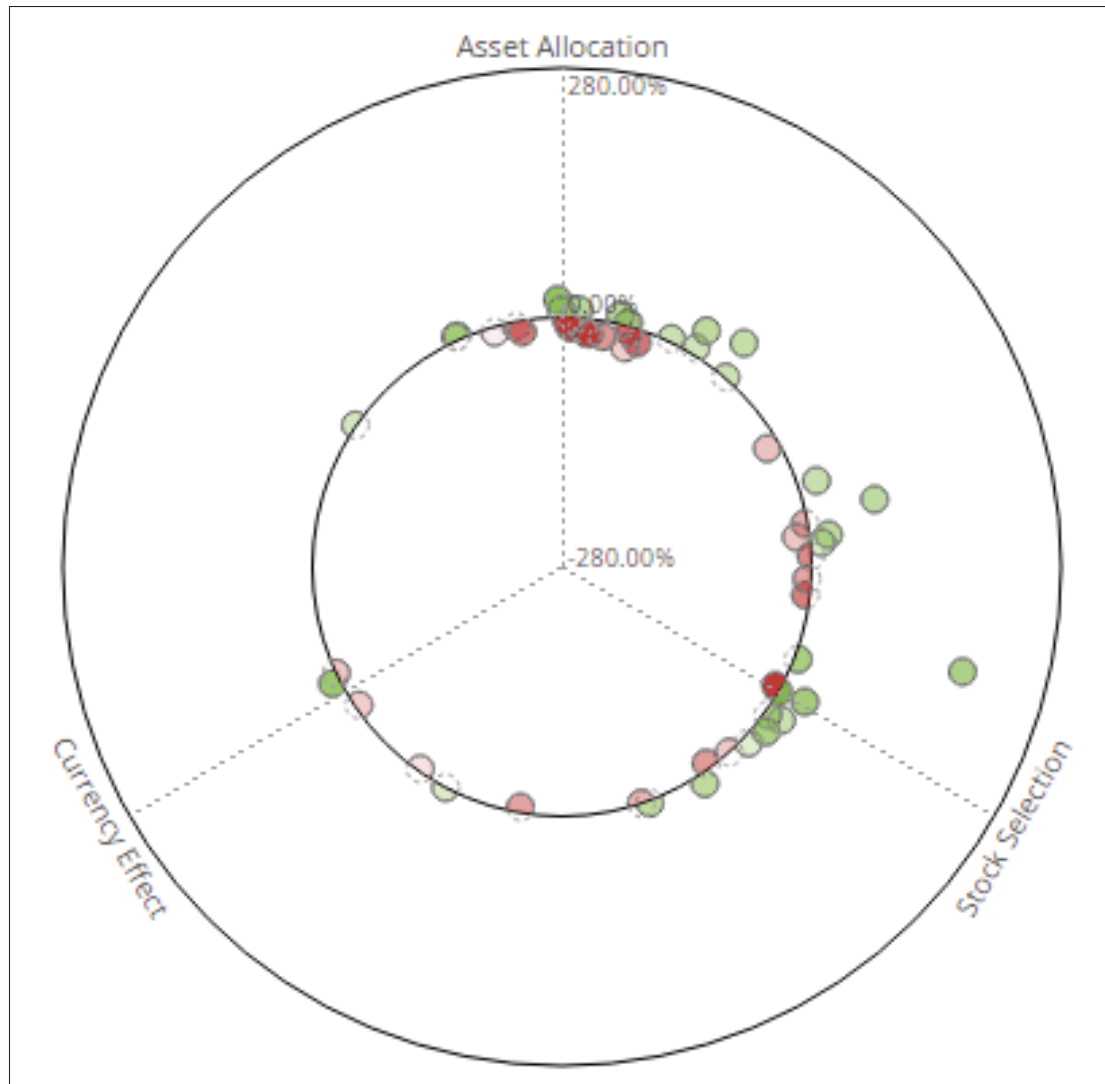
Figure 4.11   This example shows that no real strategy was applied. There is roughly as much losses as there are gains, and they are spread all over the circle. Any gain would basically be luck.

The navigation bar received more mitigated comments. Most appreciated it at the firm level when there were less data. At the advisor level, there is not enough room to display every single client. Only the biggest ones stand out. Also, there are so many clients that it is nearly impossible to conclude with color only. Even though the color mixing process is not complicated, with too many elements in the portfolio, it becomes impossible to interpret accurately. But when we asked if it should be removed or replaced, experts stated that it still had some value

and that it would probably be more useful over time after a certain learning period, and that the control could be improved with different sorting options.

Tableau 4.1   Appreciation scores for the proposed dashboard

Criteria	Score /5
Allows to easily detect problems in the dataset	4.25
Allows to explain problems easily	4.25
Easy to learn	4.5
Usability	4.75
Completeness	4.25
Overall satisfaction	4.75

4.9   Discussion

In the study, none of our experts concluded based on attribution color in the navigation bar, even in the first part of the experiment where the bar contained at most 3 elements. Concluding on attribution using only colors requires first associating mentally the effects with the colors, which can hardly be done the first time the expert uses the system. Based on our own experiences, we found that it took many days using the tool before we became accustomed to this technique.

We think experts should not be expected to conclude precisely on attribution effect for a single element in the portfolio using only the resulting color. The color encoding is not precise and is more intended for global pattern discovery. Experts confirmed that they mostly used the upper row of the navigation bar (the differential return) since it was easier for them to use the popular red-green encoding. Still, when we proposed to remove the attribution row from the bar since nobody was using it, they objected to the idea.

Based on their comments, one of the following modifications we did was to add sorting in the navigation bar. By default, elements are sorted by weight. This sorting might be useful for prioritizing analysis. But finding attribution patterns might be more difficult using this technique. In order to give better insight on the overall tendencies, we allowed sorting by differential return and by attribution effects. Sorted using these two techniques, colors follow a more linear transition and patterns emerged more easily (figure 4.12). We discussed the corrected navigation bar with our experts and they confirmed that this simple correction made the tool more useful on larger datasets. This technique is especially useful to identify the dominant effects on the global portfolio (portfolio containing all positions), since important elements in the portfolio will be represented by larger cells.
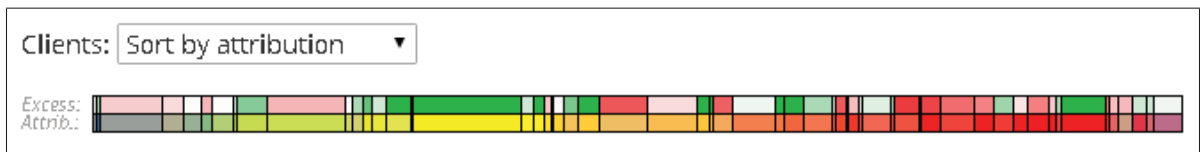
Figure 4.12   Updated navigation bar supporting data sorting by excess return and attribution effects.

Even though the attribution colored cells did not initially serve our experts, most of them liked the RYB ternary triangle displayed in the tooltip on a navigation bar's cell mouse over. Most referred either to the ternary wheel or to the ternary triangle to answer which effect dominated on a specific portfolio element. Even though none of them had previously encountered a ternary plot, they found it intuitive and easy to use.

The ternary wheel was greatly appreciated by our experts. Most found it especially useful on large datasets where it would be difficult to conclude based only on raw data. Having both the excess return and the attribution effect on the same charts give the complete picture and allows for easy pattern discovery. All experts were clearly able to conclude on advisor's strategies using the wheel and easily found outliers using the chart.

Since the ternary wheel uses a projection to reduce the dimensionality of the dataset, some of the data is abstracted. As discussed previously, the missing dimension is encoded as transparency. Since this concept is somewhat abstract, we didn't expect our experts to master this detail. Although they asked for explanations during the study, three of them actually did understand the concept and considered it when performing their analysis. For example, one expert said : "I see that this province underperformed the benchmark since it is red and under the center circle. It is between asset allocation and stock selection, but since the circle is almost transparent, it is almost neutral, currency effect is also having an important effect, although not dominant.".

Experts appreciated having access to the raw values combined with the charts. Experienced experts said that they usually only used Microsoft Excel to conclude on performance attribution. Having both the charts to guide their first reflections and the raw values to dig deeper into the data was in most cases the highlight of the system for them. And since all views were coordinated, users we able to easily link the concepts. One expert said : "Each data table presents really useful information, but we really get the big picture when we combine the views ! This is what makes this system powerful !".

When asked about the challenges of bringing this tool into a real production environment, most guessed that the learning curve might be a problem with some less tech-savvy users, but that the resulting increases in productivity would be worth it. All experts found the system easy to use (table 3.7), even though it introduced many new visualization concepts. All of them succeeded in using the tool to detect and explain real situations in the dataset on their first try. Two experts suggested that such a tool could even be used with their clients in order to explain them performance anomalies to them as it clearly illustrates the facts, and with some basic explanation, might be easier to understand than large data grids.

4.10   Conclusion

We have presented two novel visualization techniques derived from the ternary plot : the ternary wheel and the ternary navigation bar. We also developed a complete attribution analysis platform including our novel techniques and evaluated it on a real dataset of investment portfolios with four experts in portfolio management. Using our tools, experts were able to detect and explain performance anomalies at different levels.

Even though we did not intend to address the performance appraisal phase of performance analysis, it is possible to some extent to determine whether the gains and losses are due to luck or skill using the ternary wheel, as show by figures 4.10 and 4.11. A good investment strategy would show some obvious patterns with all accounts under management presenting similar effects. If a bad strategy or no strategy at all were used, we would end up with no obvious pattern, or a negative pattern with some outliers that compensate for the bad decisions. To this extent, our proposed ternary wheel enables performance attribution and performance appraisal alike. Future work could extend these concepts and cover performance appraisal more deeply from a mathematical standpoint.

Our attribution dashboard was developed as an evaluation platform for our visualization techniques and was not expected to reach the market. Based on the comments we collected from our experts, more work might be invested into this project in collaboration with our partner to develop a complete attribution solution. For example, since we evaluated our platform with only a subset of the real dataset, scalability might prove to be an issue and should be addressed in future work.

CONCLUSION GÉNÉRALE

L'objectif initial fixé par notre partenaire industriel consistait à proposer des techniques novatrices permettant de simplifier l'analyse des performances. Le principal défi pour l'entreprise : la quantité impressionnante de données impliquées. Les techniques conventionnelles ne permettent simplement pas aux experts d'obtenir une vue globale des performances et de cibler les éléments problématiques. L'analyse manuelle, principalement à l'aide de chiffriers de formules automatisées, ne permet pas d'obtenir les résultats escomptés.

Trois idées novatrices ont été explorées dans le cadre de cette thèse en lien avec cet objectif. Dans un premier temps, nous avons présenté VectorLens, une technique permettant de simplifier l'interaction avec de grands ensembles de courbes dans un graphique linéaire simple. Grâce à sa sélection angulaire, VectorLens permet de sélectionner des ensembles de courbes présentant des patrons intéressants et d'obtenir des informations complémentaires sur celles-ci. Notre approche se distingue principalement des autres techniques de sélection angulaire par le fait qu'un seul clic est nécessaire pour saisir tous les paramètres de la sélection (angle et tolérance). Une étude plus fondamentale a d'ailleurs été effectuée afin d'isoler la technique de sélection angulaire proposée par la lentille et de la comparer à l'état de l'art. Les résultats ont démontré que la sélection angulaire, combinée au pinceau (mode hybride), était la solution la plus précise et la plus appréciée des participants. La lentille VectorLens a ensuite été conçue à partir de ces résultats et intègre la sélection hybride. L'outil permet également des sélections plus complexes à l'aide de filtres catégoriques ou en combinant plusieurs lentilles afin de créer des conjonctions / disjonctions d'ensembles. L'article présenté au chapitre 2 a été publié dans le journal Transactions on Visualization and Computer Graphics en mars 2015 (Dumas *et al.*, 2015).

Le deuxième projet adresse le problème de surcharge d'information sur le même graphique. La technique proposée consiste à compresser les différentes couches d'information de façon à optimiser l'espace entre deux moments dans le temps (ticks) sur l'abscisse. La compression permet de limiter l'occlusion causée par la superposition des courbes en positionnant chaque couche de façon stratégique. Trois variantes ont été développées dans le cadre du projet pré-

senté dans le chapitre 3 : la superposition (superposed), la juxtaposition (juxtaposed) et le décalage (shifted).

Ces techniques ont été comparées à l'état de l'art dans différentes tâches où les participants devaient comparer des valeurs entre les couches, évaluer des intervalles, etc. Les résultats ont démontré que deux approches, soit juxtaposé et superposé, se démarquent particulièrement. Juxtaposé porte le flambeau lorsqu'il est question de comparer des valeurs sur l'ordonnée entre les couches. Superposé performe mieux que toutes les autres techniques lorsqu'il est question de comparer des pentes, notamment sur de grands intervalles. Il s'agit également de la technique préférée des participants pour effectuer les tâches comparées.

Enfin, le dernier projet propose deux nouvelles techniques permettant d'effectuer l'attribution des performances financières de portefeuilles. Les deux techniques proposées, la barre de navigation ternaire et la roue ternaire, sont des variations du graphique ternaire (ternary plot). Les deux techniques permettent de représenter l'excès de rendement par rapport à la référence, ainsi que la relation entre les trois effets de l'attribution, sur un même graphique. Une solution complète d'attribution, intégrant les outils proposés, a également été développée et évaluée auprès de quatre experts en finance dans un contexte d'analyse réel. Les experts ont démontré que les contributions proposées permettaient réellement de simplifier le processus d'attribution sur de grands ensembles de données. Les experts ont même été en mesure de conclure sur les grandes lignes des stratégies employées par les analystes financiers à l'aide de la roue ternaire.

Puisque les trois projets sont étroitement reliés, il ne serait pas très difficile, par exemple, d'augmenter le graphique linéaire du prototype de système d'attribution pour supporter VectorLens et les couches compressées, et ainsi créer un seul système complet d'analyse des performances intégrant l'ensemble des contributions. Cette intégration n'a pas été effectuée principalement pour des contraintes techniques, puisque les deux premiers projets ont été réalisés en C# avec WPF, alors que le dernier a été réalisé en HTML et Javascript avec D3.js (Bostock *et al.*, 2011). Il aurait donc été nécessaire de tout réimplémenter les deux premières techniques pour les intégrer. Puisque le système d'attribution visait uniquement à évaluer les variations ternaires, ces

efforts supplémentaires n'ont pas été jugés pertinents dans le cadre du projet. Cette introduction aurait également introduit un certain niveau de complexité supplémentaire qui aurait pu biaiser les résultats de l'expérience avec les experts. Toutes les techniques développées dans le cadre de ce projet sont malgré tout complémentaires et pourraient être utilisées dans un même système complet d'analyse des performances.

Une revue de la littérature en visualisation financière a également été effectuée et les résultats ont été publiés sous la forme d'une affiche lors de la conférence IEEE Vis 2014 (Dumas *et al.*, 2014). Un site web a également été réalisé (http://www.financevis.net) afin de rendre disponible les résultats du projet à toute la communauté.

4.10.1   Retombées industrielles et académiques

Le cas d'utilisation exploré dans le cadre de l'article VectorLens concerne les performances d'actions boursières de différentes industries. Puisque les cours historiques des actions sont publiquement accessibles, ces données étaient plus intéressantes pour une publication scientifique (données non confidentielles, plus facile de reproduire les résultats de l'expérience, etc). Toutefois, VectorLens a été développé avec le contexte d'analyse de portefeuilles en tête, puisqu'il s'agit du mandat donné par Croesus. Dans ce contexte, VectorLens permettrait la sélection des portefeuilles présentant des patrons similaires et de comparer les caractéristiques communes (ex : titres/secteurs détenus, conseillers, etc.). Un premier prototype a été réalisé dans le cadre du projet avec des données aléatoires, simplement pour illustrer le concept. Un prototype plus avancé est maintenant sur la planche à dessin à l'interne chez notre partenaire afin d'évaluer le concept sur de véritables données. Les gestionnaires de Croesus ont démontré un intérêt pour introduire une variante de VectorLens dans une version future de l'application.

Dans le même optique, les couches compressées ont originalement été développées dans un contexte d'analyse de performances de portefeuilles. À l'origine, la compression a été développée afin d'offrir un aperçu du niveau inférieur dans la hiérarchie des portefeuilles présentée à la section 1.1.2. Il s'agissait alors d'une technique focus+context, permettant d'afficher les

détails de la composition d'un portefeuille en surbrillance en préservant le contexte d'origine. Il s'agit d'ailleurs, pour Croesus, de l'utilisation la plus probable de la technique. Ce concept devrait d'ailleurs être intégré au même prototype que pour VectorLens, puisqu'il s'agit du même graphique sous-jacent.

Comparer la composition de plusieurs portefeuilles à l'aide de la compression pourrait toutefois également être un avantage intéressant pour Croesus. Cette idée s'intégrerait possiblement mieux dans un outil comme celui présenté dans le cadre du projet du chapitre 4.

Le dernier projet, lié à l'attribution des performances, un peu plus concret que les deux premiers, a particulièrement intéressé la direction de Croesus. Des évaluations plus poussées devraient être effectuées auprès de clients de la firme afin d'évaluer l'intérêt de la solution dans un contexte plus commercial. Des modifications importantes devront évidemment être apportées afin de mieux supporter toutes les règles d'affaires et s'intégrer un peu mieux dans l'écosystème de Croesus.

Les retombées dépassent évidemment les frontières de Croesus. Par exemple, le site http://www.financevis.net tente de connecter la communauté de visualisation financière en leur offrant un endroit commun où retrouver les dernières innovations dans le domaine. De plus, bien que les différents projets de cette thèse ont tous été développés dans un contexte d'analyse de performances, les solutions proposées sont suffisamment génériques pour être réutilisées dans d'autres domaines de la visualisation. Par exemple, une analyse préliminaire a été lancée pour modifier le panneau de contrôle développé pour l'attribution de performances afin de lui permettre d'analyser la composition de fonds mutuels et d'autres fonds échangés en bourse. VectorLens et les couches compressées pourront être réutilisés dans d'autres sphères de la visualisation (à l'extérieur du domaine de la finance), où les données présenteront une structure similaire à celles du projet initial.

L'intérêt des trois projets dépasse également les frontières de la finance. Bien qu'ils aient été créés avec un concept financier en tête, ils s'adaptent très bien à d'autres réalités. Par exemple, le concept de sélection angulaire peut être utilisé aussi bien pour analyser des séries temporelles

que sur des coordonnées parallèles. Nous avons même proposé une extension permettant de supporter des courbes paramétriques. Les couches compressées permettent de représenter toute forme de données structurées en arborescence (focus+context) ou en groupes. Enfin, les deux variations des graphiques ternaires fonctionnent dans tous les cas où des graphiques ternaires peuvent être utilisés.

ANNEXE I

FINANCEVIS.NET : A VISUAL SURVEY OF FINANCIAL DATA VISUALIZATIONS

Maxime DUMAS[1], Michael J. McGUFFIN[1], Victoria L. LEMIEUX[2]

[1] Département de Génie Mécanique, École de Technologie Supérieure,

1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

[2] University of British Columbia, 2329 West Mall, Vancouver, BC, Canada V6T 1Z4

1. Abstract

We surveyed more than 50 papers that present visualizations of financial data, and made their references available on a website, http ://financevis.net, that allows a user to browse thumbnails of the visualizations, and filter them by multiple criteria or perform a plain-text search. We discuss how we classified the visualizations, and outline ways to generate ideas for designing visualizations of financial data

2. Introduction

The financial industry involves complex network flows of vast sums of money. In 2014, the world market capitalization reached a new record high of $62.4 trillion (Perry, 2014), and each day, $5.3 trillion are traded in foreign exchange markets (Monetary and Economic Department, 2013). Tasks related to finance can involve the monitoring of large numbers of frequent transactions and price fluctuations, and complex decision making that is sometimes done under tight time constraints. Visualization is one tool that can help with such tasks, and is already frequently used, such as in line graphs and candlestick charts. To help organize the field of visualization applied to financial data, we have identified over 50 papers, from 1992 onward, that present over 70 visualizations of financial data, and we have tagged these papers with meta-data and made their references available on an interactive website (Figure I-1). Our work is inspired by other recent surveys of domain-specific visualizations : http://treevis.net (Schulz,

2011), http://survey.timeviz.net (Aigner *et al.*, 2011), http://setviz.net (Alsallakh *et al.*, 2014), and http://dynamicgraphs.fbeck.com (Beck *et al.*, 2014).
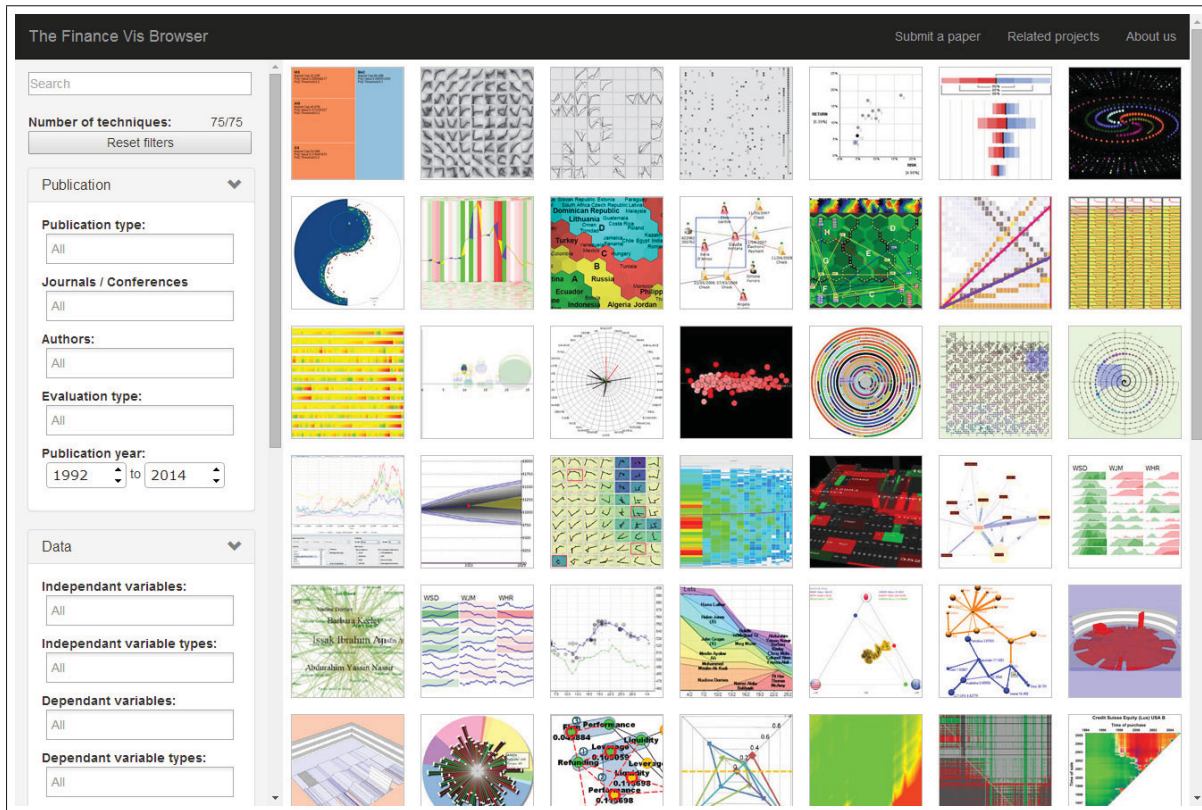
Figure-A I-1   Screenshot of our website http://financevis.net.

3.   Characterization of Papers

For each paper in our survey, we identified the one or more most interesting or novel visualizations in the paper, and catalogued these for our website. For example, if a paper presented a 3D variant of a treemap as well as a simple line graph, we included the 3D treemap in our survey, but not the simple line graph. Also, if a user interface with multiple views was presented, we only catalogued the most novel (sub)views within the interface, not the composite view.

In all of the papers surveyed, the data being visualized could be thought of as either multi-dimensional multivariate (mdmv) data (Wong et Bergeron, 1997), a tree, a network, or some

combination of these. We first discuss the mdmv aspect, and then the tree and network aspects (which we refer to as Topology on our website).

In the visualization literature, the term mdmv refers to data that could be stored as tuples, where each tuple is a row in a table whose columns are either independent or dependent variables. Some authors refer to these variables as dimensions and measures (Stolte *et al.*, 2002), respectively, and many authors distinguish between quantitative (e.g., continuous, discrete) and categorical variables. Time is often thought of as a quantitative variable, however if Time is also an independent variable, it must be somehow discretized for storage in memory. For each visualization in our survey, we noted the number (0, 1, or Multiple) of independent and dependent variables, and also the types of variables (Categorical, Quantitative, and Time as a special 3rd case to avoid confusion). Our website allows the user to filter on these, and also filter on the type of chart (bar chart, line graph, etc.).

Turning to Topology, some of the visualizations we surveyed displayed a tree or network structure in the data, often in combination with mdmv aspects of the data. For example, a treemap could be used to show hierarchical groupings of sectors that change over time, or transactions in a network occurring over time. Each visualization in the survey has a Topology of either Tree, Network, or None, which the user can filter on.

Separate from the number and type of variables in the data, and from the topology of the data, are the number of spatial variables in the visualization (catalogued as either 2D or 3D) and the use or non-use of animation in the visualization (catalogued as either Dynamic or Static, respectively). These are also available for filtering.

The tasks associated with each visualization were catalogued according to a taxonomy of financial functions and activities (line of business oriented). This taxonomy is based partly on Lemieux *et al.* (2011), and covers areas within retail banking, investment banking, asset management, and risk management. Most of the visualizations in our survey are oriented toward asset management, with a few oriented toward risk management (either financial stability monitoring, or fraud and market abuse detection and prevention). Several categories in our taxo-

nomy are not yet populated, however they are available for use as the bibliography grows, and also suggest areas for future visualization work.

Finally, in cases where papers discussed interaction techniques, we catalogued these using the categories of interaction in Yi *et al.* (2007), and also made these available for filtering the bibliography.

4. How to Choose a Visualization Technique

Designers seeking ideas for visualizing their own data can use our website in several ways. If their data has tree or network aspects, they may filter for examples with such a topology. Designers can also filter by type or number of independent or dependent variables to see pertinent examples.

In the case of mdmv data, our survey encountered bar charts, line graphs, scatter plots, heat maps, and parallel coordinates. A few techniques were more exotic, such as the "Linked triangle chart" of Suzuki *et al.* (2008), and the "Galaxy" and "Tai Chi" charts in Lei et Zhang (2011). Users may filter by visualization Type and/or Alignment (i.e., Perpendicular, Radial, Free, or Other), making it easy to quickly identify uses of, say, parallel coordinates, or the more unusual chart visualizations.

Tree structures can be visualized with node-link diagrams or with treemaps. However, in our survey, we saw treemaps used almost exclusively to show tree structures. Networks can be visualized using node-link diagrams or adjacency matrices, however we saw node-link diagrams used almost exclusively, with one interesting exception appearing in Kirkland *et al.* (1999) which used a sequence diagram to show transactions over time. The use of node-link diagrams for tree structures, and of adjacency matrices for networks, may be opportunities for financial visualization that have been underexplored.

Additional design ideas for visualizing financial data can be found in the existing surveys mentioned in the introduction, i.e., surveys of visualizations of trees, of dynamic graphs, and of time-varying data.

We have also identified a few general approaches for showing topology and simultaneously showing one or more variables (or time-varying data), some of which could yield novel results. First, consider showing topology with a node-link diagram, or treemap, or adjacency matrix. In any of these cases, dependent variables that are a function of the node can be shown as a color or glyph on each node. In the case of a treemap, the size of nodes can reflect another variable (however, this runs the risk of making certain nodes very small, which may not be desirable). Variables that are a function of node and of time can also be shown as sparklines (Tufte, 2006) on each node. Secondly, consider showing topology along one spatial axis (in the case of a network, this can be done with an arc diagram, and in the case of a tree, this can be done by laying out leaf nodes along an axis). It is then possible to show variables changing with time along a perpendicular axis. For example, Greilich *et al.* (2009) show both a tree structure and arc diagrams along one direction, and a time axis along the other direction, but not for financial data. Thirdly, it is possible to show topology within a plane, and time on a perpendicular axis, however we do not recommend such use of 3D, since having multiple 2D coordinated views avoids occlusion and navigation problems (Munzner, 2008).

5. Resources for Researchers

In addition to reviewing the literature covered by our website, readers interested in pursuing research in financial visualization can look at state-of-the-art industrial tools, including, but by no means limited to, Advanced GET by eSignal.com, MarketWatch.com's "Map of the Market" based on Wattenberg (1999), and products by OmniTrader.com.

Several data sources are also available. Examples include Yahoo Finance, Google Finance and Oanda (Forex), which offer multiple free data feeds for asset management research. Note however that they may contain a higher number of errors or missing information and might not

be the best sources for all projects. Commercial feeds, such as those provided by Bloomberg or Thomson Eikon, might be more accurate or complete. Researchers can also look to specialized distributors such as the Center for Research in Security Prices (CRSP), the Wharton Research Data Services (WRDS), the International Monetary Fund (IMF), the Bank for International Settlements or the World Bank for data.

6. Conclusion

Our website can be used to review previous literature, to answer questions about the literature (e.g., "How many papers involve Tree structures visualized in 3D ?"), to find examples related to a given data set (based on the variables involved, and/or topology, and/or task), to identify the most unusual types of visualizations, or to find combinations that have not been attempted yet (such as Network structures visualized with an adjacency matrix or an arc diagram).

Authors are invited to submit new papers to be added to our bibliography by clicking on the "Submit a paper" option on our website and filling out the necessary information. We plan to eventually prepare a survey article about this work.

ANNEXE II

AVANTAGES ET DÉSAVANTAGES DES TECHNIQUES PERMETTANT DE REPRÉSENTER PLUSIEURS COUCHES D'INFORMATION

Tableau 2.1   Advantages and disadvantages of techniques based on user comments

Technique	Advantages	Disadvantages
Conventional	– No distortion, discontinuity or shifting – Easy to learn and intuitive – Best technique for small datasets	– Occlusion – Scalability issues – Need distinctive colors for layers
Superposed	– Small shifting and discontinuity – Less occlusion than conventional – Easy to learn	– Background noise (underlaying layers) – Uneven slopes distortion – Discontinuity – Confusion with masking (see figure 3.4)
Juxtaposed	– Equal slopes distortion between layer – No occlusion – No background noise	– Larger distance between layers – Confusion between beginning / end of layer – Large distortion caused by compression – Harder to get the global overview of the data

Tableau 2.1 (suite)

Technique	Advantages	Disadvantages
Shifted	– No distortion – Partial discontinuity (because of transparency)	– Harder to master – Background noise (transparency, overlayed curves) – Gives the impression it happened later in time (because of shifting on the x-axis) – Scalability issues – Layer separation harder to see when layers are uneven
Vertical Small Multiples	– No distortion – No discontinuity – No occlusion – Easy to compare points within a single layer	– Slopes are distorted (to fit in the allocated space) – Disconnection on the Y axis (hard to compare values) – Duplicated y axes
Horizontal Small Multiples	– No distortion – No discontinuity – No occlusion – Easy to compare points within a single layer	– Slopes are distorted (to fit in the allocated space) – Disconnection on the X axis (hard to compare values) – Duplicated X axes

BIBLIOGRAPHIE

Aboelsaadat, W. et R. Balakrishnan. 2004. « An Empirical Comparison of Transparency on One and Two Layer Displays ». In *Proceedings of British HCI Group Annual Conference (HCI)*. (Leeds Metropolitan University, UK 2004), p. 53–67. Springer.

Agnew, J. R. et L. R. Szykman. 2005. « Asset Allocation and Information Overload : The Influence of Information Display, Asset Choice, and Investor Experience ». *Journal of Behavioral Finance*, vol. 6, n° 2, p. 57–70.

Aigner, W., S. Miksch, W. Müller, H. Schumann, et C. Tominski. 2007. « Visualizing time-oriented data — A systematic view ». *Computers & Graphics*, vol. 31, n° 3, p. 401–409.

Aigner, W., S. Miksch, H. Schumann, et C. Tominski, 2011. *Visualization of Time-Oriented Data*. Human-Computer Interaction Series. London, UK : Springer-Verlag, 286 p.

Albuquerque, G., T. Lowe, et M. A. Magnor. 2011. « Synthetic Generation of High-Dimensional Datasets ». *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 17, n° 12, p. 2317-2324.

Alsakran, J., Y. Zhao, et X. Zhao. 2009. « Visual Analysis of Mutual Fund Performance ». In *Proceedings of International Conference on Information Visualisation (IV)*. (Barcelona, Spain 2009), p. 252–259. IEEE Computer Society.

Alsallakh, B., L. Micallef, W. Aigner, H. Hauser, S. Miksch, et P. Rodgers. 2014. « Visualizing Sets and Set-typed Data : State-of-the-Art and Future Challenges ». In *Proceedings of Eurographics/IEEE TCVG Symposium on Visualization (EuroVis)*. (Swansea, Wales, UK 2014). IEEE.

Ayatsuka, Y., J. Rekimoto, et S. Matsuoka. 1998. « Popup vernier : a tool for sub-pixel-pitch dragging with smooth mode transition ». In *Proceedings of ACM Symposium on User Interface Software and Technology (UIST)*. (San Francisco, CA, USA 1998), p. 39–48. ACM.

Bacon, C. R., 2008. *Practical Portfolio Performance Measurement and Attribution*. éd. 2. West Sussex, England : John Wiley & Sons, 400 p.

Bailly, G., E. Lecolinet, et L. Nigay. 2007. « Quinze ans de recherche sur les menus : critères et propriétés des techniques de menus ». In *Proceedings of International Conference of the Association Francophone d'Interaction Homme-Machine (IHM)*. (Paris, France 2007), p. 119–126. ACM.

Bartram, L. et C. Ware. 2002. « Filtering and Brushing with Motion ». *Information Visualization*, vol. 1, n° 1, p. 66–79.

Baudisch, P. et C. Gutwin. 2004. « Multiblending : displaying overlapping windows simultaneously without the drawbacks of alpha blending ». In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI)*. (New York, NY, USA 2004), p. 367–374. ACM.

Beck, F., M. Burch, S. Diehl, et D. Weiskopf. 2014. « The State of the Art in Visualizing Dynamic Graphs ». In *Proceedings of Eurographics/IEEE TCVG Symposium on Visualization (EuroVis)*. (Swansea, Wales, UK 2014). IEEE Computer Society.

Bederson, B. B. 2000. « Fisheye Menus ». In *Proceedings of ACM Symposium on User Interface Software and Technology (UIST)*. (San Diego, CA, USA 2000), p. 217–225. ACM.

Bertini, E., M. Rigamonti, et D. Lalanne. 2009. « Extended Excentric Labeling ». In *Proceedings of Eurographics/IEEE TCVG Symposium on Visualization (EuroVis)*. (Berlin, Germany 2009), p. 927–934. IEEE Computer Society.

Bezerianos, A., F. Chevalier, P. Dragicevic, N. Elmqvist, et J. Fekete. 2010. « GraphDice : A System for Exploring Multivariate Social Networks ». In *Proceedings of Eurographics/IEEE TCVG Symposium on Visualization (EuroVis)*. (Bordeaux, France 2010), p. 863–872. IEEE Computer Society.

Bi, X., S.-H. Bae, et R. Balakrishnan. 2010. « Effects of Interior Bezels of Tiled-monitor Large Displays on Visual Search, Tunnel Steering, and Target Selection ». In *Proceedings of Conference on Human Factors in Computing Systems (CHI)*. (Altanta, Georgia, USA 2010), p. 65–74. ACM.

Bier, E. A., M. C. Stone, K. Pier, W. Buxton, et T. D. DeRose. 1993. « Toolglass and Magic Lenses : The See-Through Interface ». In *Proceedings of Annual ACM SIGGRAPH Symposium on User Interface Software and Technology (SIGGRAPH)*. (Anaheim, CA, USA 1993), p. 73–80. ACM.

Bostock, M., V. Ogievetsky, et J. Heer. 2011. « D3 : Data-Driven Documents ». In *Proceedings of IEEE Symposium on Information Visualization (InfoVis)*. (Providence, RI, USA 2011), p. 2301–2309. IEEE Computer Society.

Bremm, S., T. von Landesberger, M. HeÃŸ, et D. Fellner. 2012. « PCDC - On the Highway to Data - A Tool for the Fast Generation of Large Synthetic Data Sets ». In *Proceedings of International Eurovis Workshop on Visual Analytics (EuroVA)*. (Bordeaux, France 2012), p. 7–11. Eurographics Association.

Buono, P. et A. L. Simeone. 2008. « Interactive shape specification for pattern search in time series ». In *Proceedings of ACM Advanced Visual Interfaces (AVI)*. (Napoli, Italy 2008), p. 480–481. ACM.

Card, S. K., J. D. Mackinlay, et B. Schneiderman, 1999. *Readings in Information Visualization : Using Vision to Think*. Interactive Technologies. London, UK : Morgan Kaufmann, 712 p.

Chapuis, O., J.-B. Labrune, et E. Pietriga. 2009. « DynaSpot : speed-dependent area cursor ». In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI)*. (New York, NY, USA 2009), p. 1391–1400. ACM.

Chewning Jr, E. G. et A. M. Harrell. 1990. « The effect of information load on decision makers' cue utilization levels and decision quality in a financial distress decision task ». *Accounting, Organizations and Society*, vol. 15, n° 6, p. 527–542.

Cockburn, A., A. Karlson, et B. B. Bederson. 2008. « A review of overview+detail, zooming, and focus+context interfaces ». *ACM Computing Surveys (CSUR)*, vol. 41, n° 1, p. 2 :1-2 :31.

Cockburn, A., A. Karlson, et B. B. Bederson. 2009. « A review of overview+detail, zooming, and focus+context interfaces ». *ACM Computing Surveys*, vol. 41, n° 1, p. 2 :1–2 :31.

Collins, C. et S. Carpendale. 2007. « VisLink : Revealing relationships amongst visualizations ». In *Proceedings of IEEE Symposium on Information Visualization (InfoVis)*. (Sacramento, CA, USA 2007), p. 1192–1199. IEEE Computer Society.

De Finetti, B. 1926. « Considerazioni matematiche sull'ereditarietà mendeliana ». *Metron - International Journal of Statistics*, vol. 6, n° 1, p. 3–41.

Driver, J., P. McLeod, et Z. Dienes. 1992. « Motion coherence and conjunction search : Implications for guided search theory ». *Perception & Psychophysics*, vol. 51, n° 1, p. 79–85.

Dumas, M., M. J. McGuffin, et V. L. Lemieux. 2014. « Financevis. net : A Visual Survey of Financial Data Visualizations ». In *Poster Abstracts of IEEE VIS 2014*. (Paris, France 2014). IEEE Computer Society.

Dumas, M., M. J. McGuffin, et P. Chassé. 2015. « VectorLens : Angular Selection of Curves within 2D Dense Visualizations ». *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 21, n° 3, p. 402-412.

D'Urso, P., C. Cappelli, D. Di Lallo, et R. Massari. 2013. « Clustering of financial time series ». *Physica A : Statistical Mechanics and its Applications*, vol. 392, n° 9, p. 2114–2129.

Dwyer, T. 2003. « A Scalable Method for Visualising Changes in Portfolio Data ». In *Proceedings of Australasian Symposium on Information Visualisation (InVis.au)*. (Adelaide, Australia 2003), p. 17–25. Australian Computer Society, Inc.

Dwyer, T. et D. R. Gallagher. 2004. « Visualising changes in fund manager holdings in two and a half-dimensions ». *Information Visualization*, vol. 3, n° 4, p. 227–244.

Economides, N. 2004. « The Impact of the Internet on Financial Markets ». *Journal of Financial Transformation*, vol. 1, n° 1, p. 8–13.

Eklund, T., B. Back, H. Vanharanta, et A. Visa. 2003. « Using the Self-Organizing Map as a Visualization Tool in Financial Benchmarking ». *Information Visualization*, vol. 2, n° 3, p. 171–181.

Ellis, G. et A. Dix. 2006. « Enabling Automatic Clutter Reduction in Parallel Coordinate Plots ». *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 12, n° 5, p. 717–724.

Elmqvist, N. et P. Tsigas. 2008. « A taxonomy of 3D occlusion management for visualization ». *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 14, n° 5, p. 1095–1109.

Eppler, M. J. et J. Mengis. 2004. « The Concept of Information Overload : A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines ». *The Information Society*, vol. 20, n° 5, p. 325–344.

Fekete, J.-D. et C. Plaisant. 1999. « Excentric labeling : dynamic neighborhood labeling for data visualization ». In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI)*. (Pittsburgh, PA, USA 1999), p. 512–519. ACM.

Few, S. 2006. « Bullet graph design specification ». <http://www.perceptualedge.com>. Consulté le 2 juillet 2015.

Findlater, L., A. Jansen, K. Shinohara, M. Dixon, P. Kamb, J. Rakita, et J. O. Wobbrock. 2010. « Enhanced area cursors : reducing fine pointing demands for people with motor impairments ». In *Proceedings of ACM Symposium on User Interface Software and Technology (UIST)*. (New York, NY, USA 2010), p. 153–162. ACM.

Giller, V., M. Tscheligi, J. Schrammel, P. Fröhlich, B. Rabl, R. Kosara, S. Miksch, et H. Hauser. 2001. *Experimental evaluation of semantic depth of field, a preattentive method for focus+context visualization*. Technical Report TR-2001-3. Vienna, Austria : Vienna University of Technology, 10 p.

Google. 2013. « Google Finance : stock market quotes ». <https://www.google.com/finance?q=msft>. Consulté le 14 mars 2013.

Greilich, M., M. Burch, et S. Diehl. 2009. « Visualizing the Evolution of Compound Digraphs with TimeArcTrees ». *Computer Graphics Forum*, vol. 28, n° 3, p. 975–982.

Grossman, T. et R. Balakrishnan. 2005. « The bubble cursor : enhancing target acquisition by dynamic resizing of the cursor's activation area ». In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI)*. (Portland, OR, USA 2005), p. 281–290. ACM.

Guimbretière, F. et T. Winograd. 2000. « FlowMenu : Combining Command, Text, and Data Entry ». In *Proceedings of ACM Symposium on User Interface Software and Technology (UIST)*. (San Diego, CA, USA 2000), p. 213–216. ACM.

Guo, P., H. Xiao, Z. Wang, et X. Yuan. 2010. « Interactive local clustering operations for high dimensional data in parallel coordinates ». In *Proceedings of IEEE Pacific Visualization Symposium (PacificVis)*. (Taipei, Taiwan 2010), p. 97–104. IEEE Computer Society.

Harrison, B. L. et K. J. Vicente. 1996. « An Experimental Evaluation of Transparent Menu Usage ». In *Proceedings of Conference on Human Factors in Computing Systems (CHI)*. (Vancouver, British Columbia, Canada 1996), p. 391–398. ACM.

Harrison, B. L., G. Kurtenbach, et K. J. Vicente. 1995. « An Experimental Evaluation of Transparent User Interface Tools and Information Content ». In *Proceedings of ACM Symposium on User Interface Software and Technology (UIST)*. (Pittsburgh, PA, USA 1995), p. 81–90. ACM.

Hauser, H., F. Ledermann, et H. Doleisch. 2002. « Angular brushing of extended parallel co-ordinates ». In *Proceedings of IEEE Symposium on Information Visualization (InfoVis)*. (Boston, MA, USA 2002), p. 127–130. IEEE Computer Society.

Heer, J. et M. Agrawala. 2006. « Multi-Scale Banking to 45 Degrees ». *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 12, n° 5, p. 701–708.

Henry Riche, N., T. Dwyer, B. Lee, et S. Carpendale. 2012. « Exploring the Design Space of Interactive Link Curvature in Network Diagrams ». In *Proceedings of International Working Conference on Advanced Visual Interfaces (AVI)*. (Capri Island (Naples), Italy 2012), p. 506–513. ACM.

Hochheiser, H. et B. Shneiderman. 2004. « Dynamic query tools for time series data sets : timebox widgets for interactive exploration ». *Information Visualization*, vol. 3, n° 1, p. 1–18.

Hoeber, O. 2009. « User Evaluation Methods for Visual Web Search Interfaces ». In *Proceedings of International Conference on Information Visualisation (IV)*. (Barcelona, Spain 2009), p. 139–145. IEEE Computer Society.

Holten, D. et J. J. van Wijk. 2009. « Force-Directed Edge Bundling for Graph Visualization ». In *Proceedings of Eurographics/IEEE TCVG Symposium on Visualization (EuroVis)*. (Berlin, Germany 2009), p. 983–990. IEEE Computer Society.

Holz, C. et S. Feiner. 2009. « Relaxed selection techniques for querying time-series graphs ». In *Proceedings of ACM Symposium on User Interface Software and Technology (UIST)*. (Victoria, BC, Canada 2009), p. 213–222. ACM.

Impactopia. 2013. « Stock corrolation ». <http://www.impactopia.com/index.php?option=com_impactopia&view=friend&Itemid=2>. Consulté le 14 mars 2013.

Investments, G. 2013. « Introduction to G-Sphere ». <http://www.gravityinvestments.com/introduction.php>. Consulté le 14 mars 2013.

Ishak, E. W. et S. K. Feiner. 2004. « Interacting with Hidden Content Using Content-aware Free-space Transparency ». In *Proceedings of ACM Symposium on User Interface Software and Technology (UIST)*. (Santa Fe, New Mexico 2004), p. 189–192. ACM.

Karnosky, D. S. et B. D. Singer. 1994. « Global Asset Management and Performance Attribution ». *Research Foundation Publications*, vol. 1994, n° 3, p. 1–75.

Keim, D., F. Mansmann, J. Schneidewind, et H. Ziegler. 2006. « Challenges in Visual Data Analysis ». In *Proceedings of International Conference on Information Visualization (IV)*. (London, England, UK 2006), p. 9–16. IEEE Computer Society.

Kincaid, R. 2010. « SignalLens : Focus+Context Applied to Electronic Time Series ». *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 16, n° 6, p. 900–907.

Kincaid, R. et H. Lam. 2006. « Line graph explorer : scalable display of line graphs using Focus+Context ». In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI)*. (Venezia, Italy 2006), p. 404–411. ACM.

Kirkland, J. D., T. E. Senator, J. J. Hayden, T. Dybala, H. G. Goldberg, et P. Shyr. 1999. « The NASD Regulation Advanced-Detection System (ADS) ». *AI Magazine*, vol. 20, n° 1, p. 55.

Kirkpatrick, C. D. et J. R. Dahlquist, 2010. *Technical Analysis : The Complete Resource for Financial Market Technicians, Second Edition*. éd. 2. Upper Saddle River, NJ, USA : FT Press, 704 p.

Kosara, R. 2011. « Indirect multi-touch interaction for brushing in parallel coordinates ». In *Proc. SPIE 7868, Visualization and Data Analysis*. (San Francisco, CA, USA 2011), p. 786809.1–786809.7. SPIE.

Kosara, R., S. Miksch, et H. Hauser. 2001. « Semantic depth of field ». In *Proceedings of IEEE Symposium on Information Visualization (InfoVis)*. (San Diego, CA, USA 2001), p. 97–104. IEEE Computer Society.

Kosara, R., C. G. Healey, V. Interrante, D. H. Laidlaw, et C. Ware. 2003. « User Studies : Why, How, and When ? ». *IEEE Computer Graphics and Applications*, vol. 23, n° 4, p. 20–25.

Kurtenbach, G. et W. Buxton. 1993. « The Limits of Expert Performance Using Hierarchic Marking Menus ». In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI)*. (Amsterdam, Netherlands 1993), p. 482–487. ACM.

Lansiluoto, A., T. Eklund, B. Back, H. Vanharanta, et A. Visa. 2004. « Industry-specific cycles and companies' financial performance comparison using self-organizing maps ». *Benchmarking : An International Journal*, vol. 11, n° 3, p. 267–286.

Lawton, P. et T. Jankowski, 2009. *Investment Performance Measurement : Evaluating and Presenting Results*. CFA Investment Perspectives. Hoboken, New Jersey : Wiley, 984 p.

Lei, S. T. et K. Zhang. 2011. « Visual signatures for financial time series ». In *Proc. of International Symposium on Visual Information Communication and Interaction (VINCI)*. (Hong Kong 2011), p. 16 :1–16 :10. ACM.

Lemieux, V. L., D. Long, et D. Kemp. 2011. Global financial markets. Coleman, L., Victoria L. Lemieux, Rod Stone, et Geoffrey Yeo, editors, *Managing Records in Global Financial Markets : Ensuring Compliance and Mitigating Risk*, p. 1–14. Facet Publishing, London.

Lemieux, V. L., B. Fisher, et T. Dang. 2014. The Visual Analysis of Financial Data. Brose, M. S., Mark D. Flood, Dilip Krishna, et Bill Nichols, editors, *Handbook of Financial Data and Risk Information II*, p. 284–331. Cambridge University Press, Cambridge, UK.

Lieberman, H. 1997. « A multi-scale, multi-layer, translucent virtual space ». In *Proceedings of First International Conference on Information Visualisation (IV)*. (London, England 1997), p. 124–131. IEEE Computer Society.

Lieberman, H. 1994. « Powers of Ten Thousand : Navigating in Large Information Spaces ». In *Proceedings of ACM Symposium on User Interface Software and Technology (UIST)*. (Marina del Rey, CA, USA 1994), p. 15–16. ACM.

Lopez-Hernandez, R., D. Guilmaine, M. McGuffin, et L. Barford. 2010. « A layer-oriented interface for visualizing time-series data from oscilloscopes ». In *Proceedings of IEEE Pacific Visualization Symposium (PacificVis)*. (Taipei, Taiwan 2010), p. 41–48. IEEE Computer Society.

Marghescu, D. 2007. *Multidimensional Data Visualization Techniques for Financial Performance Data : A Review*. Technical Report 810. TURKU, Finland : Turku Centre for Computer Science, 32 p.

McCormick, B. H., Thomas A. DeFanti, et M. D. Brown. 1987. « Visualization in Scientific Computing ». *ACM SIGBIO Newsletter*, vol. 21, n° 1, p. 15-21.

McGuffin, M., N. Burtnyk, et G. Kurtenbach. 2002. « FaST Sliders : Integrating Marking Menus and the Adjustment of Continuous Values ». In *Proceedings of Graphics Interface (GI)*. (Calgary, AB, Canada 2002), p. 35–41.

Monetary and Economic Department. 2013. *Foreign exchange turnover in April 2013 : preliminary global results*. Technical report. Basel, Switzerland : Bank for International Settlements, 24 p.

Muigg, P., J. Kehrer, S. Oeltze, H. Piringer, H. Doleisch, B. Preim, et H. Hauser. 2008. « A Four-level Focus+Context Approach to Interactive Visual Analysis of Temporal Features in Large Scientific Data ». In *Proceedings of Eurographics/IEEE TCVG Symposium on Visualization (EuroVis)*. (Eindhoven, Netherlands 2008), p. 775–782. IEEE Computer Society.

Munzner, T. 2008. Process and pitfalls in writing information visualization research papers. Kerren, A., John T. Stasko, Jean-Daniel Fekete, et Chris North, editors, *Information Visualization : Human-Centered Issues and Perspectives*, p. 134–153. Springer.

Munzner, T., 2015. *Visualization Analysis and Design*. AK Peters Visualization Series. Boca Raton, FL, USA : AK Peters / CRC Press, 428 p.

Nakayama et G. H. Silverman. 1986. « Serial and parallel processing of visual feature conjunctions ». *Nature*, vol. 320, n° 6059, p. 264–265.

Nesbitt, K. V. et S. Barrass. 2004. « Finding Trading Patterns in Stock Market Data ». *IEEE Comput. Graph. Appl.*, vol. 24, n° 5, p. 45–55.

Nison, S., 2003. *Japanese Candlestick Charting : A Contemporary Guide to the Ancient Techniques of the Far East*. éd. 2. Paramus, NJ, USA : Prentice Hall Press, 299 p.

North, C. 2006. « Toward measuring visualization insight ». *IEEE Computer Graphics and Applications*, vol. 26, n° 3, p. 6–9.

Palmer, E. M., T. C. Clausner, et P. J. Kellman. 2008. « Enhancing Air Traffic Displays via Perceptual Cues ». *ACM Transactions on Applied Perception*, vol. 5, n° 1, p. 4 :1–4 :22.

Panagiotidis, A., H. Bosch, S. Koch, et T. Ertl. 2011. « EdgeAnalyzer : Exploratory analysis through advanced edge interaction ». In *Proceeding of Hawaii International Conference on System Sciences (HICSS)*. (Kauai, HI, USA 2011), p. 1–10. IEEE Computer Society.

Paredes, T. A. 2003. « Blinded by the Light : Information Overload and Its Consequences for Securities Regulation ». *Washington University Law Quarterly*, vol. 81, p. 417.

Perry, M. J. 2014. « Global Stock Rally : World Market Cap Reached Record High In March ». <http://seekingalpha.com/article/2142523-global-stock-rally-world-market-cap-reached-record-high-in-march>. Consulté le 18 juin 2014.

Pietriga, E., O. Bau, et C. Appert. 2010. « Representation-independent in-place magnification with sigma lenses ». *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 16, n° 3, p. 455–467.

Plaisant, C. 2004. « The challenge of information visualization evaluation ». In *Proceedings of International Working Conference on Advanced Visual Interfaces (AVI)*. (Gallipoli, Italy 2004), p. 109–116. ACM.

Pook, S., E. Lecolinet, G. Vaysseix, et E. Barillot. 2000. « Control menus : execution and control in a single interactor ». In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI)*. (The Hague, Netherlands 2000), p. 263–264. ACM.

Ramos, G., G. Robertson, M. Czerwinski, D. Tan, P. Baudisch, K. Hinckley, et M. Agrawala. 2006. « Tumble ! Splat ! Helping Users Access and Manipulate Occluded Content in 2D Drawings ». In *Proceedings of Advanced Visual Interfaces (AVI)*. (Venezia, Italy 2006), p. 428–435. ACM.

Riche, N. H., T. Dwyer, B. Lee, et S. Carpendale. 2012. « Exploring the design space of interactive link curvature in network diagrams ». In *Proceedings of International Working Conference on Advanced Visual Interfaces (AVI)*. (Capri Island, Naples, Italy 2012), p. 506–513. ACM.

Sarlin, P. et T. Eklund. 2011. « Fuzzy Clustering of the Self-Organizing Map : Some Applications on Financial Time Series ». In *Advances in Self-Organizing Maps - Proceedings of Workshop on Self-Organizing Maps (WSOM)*. (Espoo, Finland 2011), p. 40–50. Springer Berlin Heidelberg.

Savikhin, A., H. C. Lam, B. Fisher, et D. Ebert. 2009. « An Experimental Study of Financial Portfolio Selection with Visual Analytics for Decision Support ». In *Proceedings of Hawaii International Conference on System Sciences (HICSS)*. (Waikoloa, Big Island, Hawaii, USA 2009), p. 1–10. IEEE Computer Society.

Schaefer, M., F. Wanner, R. Kahl, L. Zhang, T. Schreck, et D. A. Keim. 2011. « A Novel Explorative Visualization Tool for Financial Time Series Data Analysis ». In *Proceedings of International UKVAC Workshop on Visual Analytics (VAW)*. (London, UK 2011).

Schreck, T., T. Tekušová, J. Kohlhammer, et D. Fellner. 2007. « Trajectory-based visual analysis of large financial time series data ». *ACM SIGKDD Explorations Newsletter*, vol. 9, n° 2, p. 30–37.

Schulz, H.-J. 2011. « Treevis.net : A Tree Visualization Reference ». *IEEE Computer Graphics and Applications*, vol. 31, n° 6, p. 11–15.

Shneiderman, B. 1996. « The eyes have it : a task by data type taxonomy for information visualizations ». In *Proceedings of IEEE Symposium on Visual Languages (VL)*. (Boulder, CO, USA 1996), p. 336–343. IEEE Computer Society.

SmartMoney. 2013. « SmartMoney Map of the Market ». <http://www.smartmoney.com/map-of-the-market/>. Consulté le 14 mars 2013.

Spence, R., 2006. *Information Visualization : Design for Interaction*. éd. 2. Essex, UK : Prentice Hall, 304 p.

Stolte, C., D. Tang, et P. Hanrahan. 2002. « Polaris : A System for Query, Analysis, and Visualization of Multidimensional Relational databases ». *IEEE TVCG*, vol. 8, n° 1, p. 52–65.

Streit, M., H. Schulz, A. Lex, D. Schmalstieg, et H. Schumann. 2012. « Model-Driven Design for the Visual Analysis of Heterogeneous Data ». *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, n° 6, p. 998–1010.

Suzuki, Y., K. Misue, et J. Tanaka. 2008. « Visualization Technique for Linked Changes in Exchange Rates Between Three Currencies ». In *Proceedings of International Conference on Knowledge, Information and Creativity Support Systems (KICSS)*. (Hanoi, Vietnam 2008), p. 95–102.

Thomas, J. J. et K. A. Cook, 2005. *Illuminating the Path : The Research and Development Agenda for Visual Analytics*. Washington, DC, USA : IEEE Computer Society Press, 186 p.

Tominski, C., S. Gladisch, U. Kister, et H. Schumann. 2014. « A Survey on Interactive Lenses in Visualization ». In *Proceedings of Eurographics/IEEE TCVG Symposium on Visualization (EuroVis)*. (Swansea, UK 2014), p. 43–62. IEEE Computer Society.

Tominski, C. et W. Aigner. 2013. « The TimeViz Browser ». <http://www.timeviz.net/>. Consulté le 8 juillet 2015.

Tory, M. et T. Möller. 2004. « Human factors in visualization research ». *IEEE Transactions on Visualization and Computer Graphics*, vol. 10, n° 1, p. 72–84.

Tufte, E. R., 2001. *The Visual Display of Quantitative Information*. éd. 2nd. Cheshire, CT, USA : Graphics Press, 197 p.

Tufte, E. R., 2006. *Beautiful Evidence*. Cheshire, CT, USA : Graphics Press, 213 p.

van den Elzen, S. et J. J. van Wijk. 2014. « Multivariate network exploration and presentation : From detail to overview via selections and aggregations ». *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 20, n° 12, p. 2310–2319.

van Ham, F. et A. Perer. 2009. « Search, Show Context, Expand on Demand : Supporting Large Graph Exploration with Degree-of-Interest ». *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, n° 6, p. 953–960.

Wang Baldonado, M. Q., A. Woodruff, et A. Kuchinsky. 2000. « Guidelines for Using Multiple Views in Information Visualization ». In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI)*. (Palermo, Italy 2000), p. 110–119. ACM.

Ware, C., 2013. *Information Visualization : Perception for Design*. éd. 3rd. Waltham, MA, USA : Morgan Kaufmann Publishers, 512 p.

Wattenberg, M. 1999. « Visualizing the stock market ». In *Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems*. (Pittsburg, PA, USA 1999), p. 188–189. ACM.

Wattenberg, M. 2001. « Sketching a graph to query a time-series database ». In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI)*. (Seattle, WA, USA 2001), p. 381–382. ACM.

Weisbuch, G., A. Kirman, et D. Herreiner. 2000. « Market organisation and trading relationships ». *The Economic Journal*, vol. 110, n° 463, p. 411–436.

Wong, N., S. Carpendale, et S. Greenberg. 2003. « Edgelens : an interactive method for managing edge congestion in graphs ». In *Proceedings of IEEE Symposium on Information Visualization (InfoVis)*. (Seattle, WA, USA 2003), p. 51–58. IEEE Computer Society.

Wong, N. et S. Carpendale. 2007. « Supporting interactive graph exploration with edge plucking ». In *Proceedings of Visualization and Data Analysis (VDA)*. (San Jose, CA, USA 2007). SPIE.

Wong, P. C. et R. D. Bergeron. 1997. 30 years of multidimensional multivariate visualization. Nielson, G. M., Hans Hagen, et Heinrich Müller, editors, *Scientific Visualization : Overviews, Methodologies, and Techniques*, p. 3-33. IEEE Computer Society, Washington, DC, USA.

Yatani, K., K. Partridge, M. Bern, et M. W. Newman. 2008. « Escape : a target selection technique using visually-cued gestures ». In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI)*. (Florence, Italy 2008), p. 285–294. ACM.

Yi, J. S., Y. a. Kang, J. Stasko, et J. Jacko. 2007. « Toward a Deeper Understanding of the Role of Interaction in Information Visualization ». *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, n° 6, p. 1224–1231.

Zhao, J., F. Chevalier, E. Pietriga, et R. Balakrishnan. 2011a. « Exploratory Analysis of Time-Series with ChronoLenses ». *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 17, n° 12, p. 2422–2431.

Zhao, J., F. Chevalier, et R. Balakrishnan. 2011b. « KronoMiner : using multi-foci navigation for the visual exploration of time-series data ». In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI)*. (Vancouver, BC, Canada 2011), p. 1737–1746. ACM.

Ziegler, H., M. Jenny, T. Gruse, et D. Keim. 2010. « Visual market sector analysis for financial time series data ». In *Proceedings of IEEE Symposium on Visual Analytics Science and Technology (VAST)*. (Salt Lake City, Utah, USA 2010), p. 83–90. IEEE Computer Society.

Ziegler, H., T. Nietzschmann, et D. A. Keim. 2008. « Visual Analytics on the Financial Market : Pixel-based Analysis and Comparison of Long-Term Investments ». In *Proceedings of International Conference on Information Visualisation (IV)*. (London, UK 2008), p. 287–295. IEEE Computer Society.