

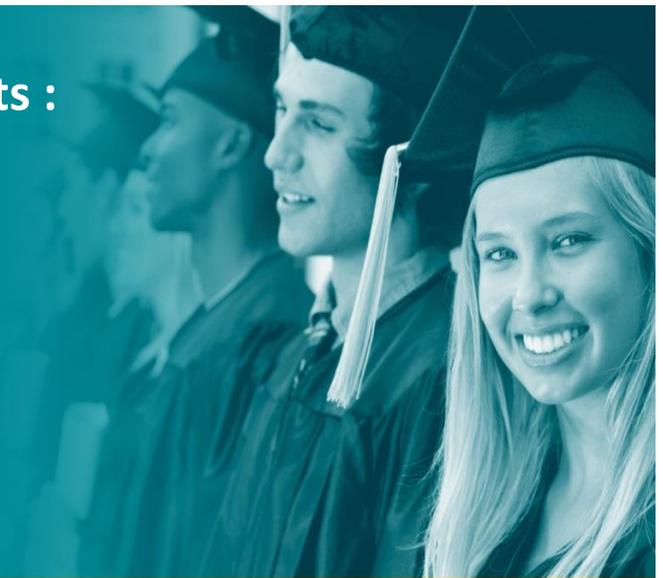


Conseil ontarien
de la qualité de
l'enseignement
supérieur

Un organisme du gouvernement de l'Ontario

Visualisation des renseignements : Guide de l'utilisateur relatif aux données sur l'éducation

Ajay Sivanand et Brian Frank
Université Queen's



Publié par

Le Conseil ontarien de la qualité de l'enseignement supérieur

1, rue Yonge, bureau 2402
Toronto (Ontario) Canada, M5E 1E5

Téléphone : 416 212-3893
Télécopie : 416 212-3899
Site Web : www.heqco.ca/fr-CA/
Courriel : info@coqes.ca

Se référer au présent document comme suit :

Sivanand, A. et Frank, B. (2018). *Visualisation des renseignements : Guide de l'utilisateur relatif aux données sur l'éducation*. Toronto : Conseil ontarien de la qualité de l'enseignement supérieur.



Les opinions exprimées dans le présent rapport de recherche sont celles des auteurs et ne reflètent pas nécessairement le point de vue ni les politiques officielles du Conseil ontarien de la qualité de l'enseignement supérieur ou des autres organismes ou organisations ayant offert leur soutien, financier ou autre, dans le cadre de ce projet. © Imprimeur de la Reine pour l'Ontario, 2018

Table des matières

Visualisation des renseignements : Guide de l'utilisateur relatif aux données sur l'éducation.....	1
Table des matières.....	2
Introduction.....	5
Guide de conception de la visualisation.....	6
Étape 1 : Comprendre le problème.....	7
<i>Abstraction des données : Quelles données sont présentées?</i>	7
<i>Abstraction de la tâche : Pourquoi est-ce illustré?</i>	11
Étape 2 : Attribuer des attributs.....	14
Étape 3 : Peaufiner la visualisation.....	18
Limites du guide et pratiques exemplaires.....	18
Exemples du Guide.....	19
Exemple 1 : « Les groupes sous-représentés présentent-ils le même taux de diplomation que les groupes fortement représentés? ».....	19
<i>Étape 1 : Abstraction des données et des tâches</i>	19
<i>Étape 2 : Attribuer des attributs</i>	20
<i>Étape 3 : Peaufiner la visualisation</i>	21
Exemple 2 : « Quels sont les cours les plus fréquemment choisis pour les étudiants d'un groupe démographique en particulier? ».....	22
<i>Étape 1 : Abstraction des données et des tâches</i>	22
<i>Étape 2 : Attribuer des attributs</i>	22
<i>Étape 3 : Peaufiner la visualisation</i>	23
Références.....	26
Glossaire.....	27

Liste des tableaux

Tableau 1 : Résumé des catégories et des options d'abstraction des données	7
Tableau 2 : Liste des types d'attribut avec exemples	10
Tableau 3 : Sommaire des niveaux d'abstraction des actions et des options correspondantes pour chaque action	12
Tableau 4 : Résumé des types de cibles et de l'état des tâches auxquels ils peuvent s'appliquer	13
Tableau 5 : Liste ordonnée des voies (adaptée de la figure 5.1 dans Munzner, 2014)	15
Tableau 6 : Tableau de données abstraites pour une question sur les taux de diplomation.....	20
Tableau 7: Tableau de données abstraites pour la question sur le choix de cours	22

Liste des figures

Figure 1 : Exemples d'éléments et d'attributs en format tabulaire.....	8
Figure 2 : Exemples de comparaison de groupes semblables dans un ensemble de données	17
Figure 3 : Visualisation de « Les groupes sous-représentés présentent-ils le même taux de diplomation que les groupes fortement représentés? »	21
Figure 4 : Première version de la visualisation pour « Quels sont les cours les plus fréquemment choisis pour les étudiants d'un groupe démographique en particulier? »	23
Figure 5 : Version mise à jour de la visualisation pour « Quels sont les cours les plus fréquemment choisis pour les étudiants d'un groupe démographique particulier? »	24

Remerciements

Ce rapport n'aurait pas pu être de la qualité, sans l'aide de Jake Kaupp et de David Waller. Merci à Gregory Hicks pour son expérience en rédaction technique et pour son aide dans la structure et la clarté du rapport.

Introduction

Dans le cadre de l'interprétation ou de la communication de renseignements avec les données, les visualisations de renseignements jouent un rôle prédominant, voire peut-être crucial. Les outils de visualisation des données et les bibliothèques comme un [graphique brut](#) et des [catalogues de visualisation de données](#) peuvent aider les utilisateurs à visualiser un ensemble de données. Compte tenu des nombreux outils disponibles, il importe de savoir quelles représentations visuelles peuvent être exécutées le plus efficacement possible tout en ayant un maximum d'utilité.

Les membres du corps professoral et du personnel de niveau postsecondaire qui cherchent à accroître leur utilisation des données sur l'éducation pour orienter la prise de décisions font régulièrement face aux limites des outils et approches traditionnels d'analyse des données. Les ensembles de données et les questions auxquelles ils doivent répondre se sont élargis et sont devenus plus complexes. Il est donc plus difficile de générer des idées utilisables au moyen des méthodes conventionnelles. Peu d'outils logiciels peuvent servir à intégrer des données pédagogiques complexes aux décisions d'amélioration des programmes du point de vue des cours, des départements ou des établissements. Une liste complète de ces outils se trouve dans notre examen des travaux en cours dans le domaine (Sivanand et Frank, 2017).

Ce guide a été créé dans le cadre d'un projet de recherche plus vaste qui avait pour but de trouver des façons de promouvoir l'utilisation des données pédagogiques au niveau postsecondaire au moyen de visualisations de renseignements. Il vise à aider les membres du corps professoral et du personnel de soutien à l'éducation à concevoir davantage de visualisations enrichies de renseignements qui permettent de mieux utiliser les données sur l'éducation qui sont disponibles et de donner un aperçu des domaines d'enquête qui sont plus complexes.

Le projet de recherche a identifié les principaux intervenants impliqués dans le processus de création de la visualisation ainsi que les principaux défis auxquels ils font face actuellement. Une application Web en ligne appelée [Easel: Visualization Recommendation System \(VRS\)](#) a été élaborée pour appuyer plusieurs de ces défis, plus particulièrement ceux qui consistent à déterminer les deux questions à poser au sujet des données et les façons de répondre à ces questions. Ces questions comprennent : « Comment les étudiants d'autres départements réussissent-ils dans les cours que nous offrons? » et « Les notes aux questionnaires permettent-elles de prédire la réussite à une question semblable à l'examen final? » Les visualisations recommandées ont été conçues en fonction de l'expertise acquise au moyen de diverses ressources (Tuft et Graves-Morris, 1983; Bertin, 1981; Börner et Polley, 2014), mais principalement le manuel de Munzner portant sur l'analyse et la conceptualisation de la visualisation (2014). Sivanand donne tous les détails du grand projet de recherche et du développement d'Easel (2017).

L'intention du système de recommandation en matière de visualisation d'Easel n'a jamais été de formuler des recommandations pour toutes les questions qu'un utilisateur pourrait avoir. Nous avons résumé les sujets pertinents du manuel de Munzner en une méthodologie simplifiée afin de créer des visualisations conçues spécialement pour être utilisées avec les données éducatives. Le présent guide vise à aider une

personne qui souhaite se servir de ses données sur l'éducation obtenues à la suite d'une interrogation et les illustrer à titre de visualisation qui répondrait le mieux à ses besoins.

Conformément à l'approche de Munzner en matière de conception de visualisation, le présent guide met l'accent sur l'abstraction du type de données et des tâches à accomplir. Le processus d'abstraction consiste à faire la transition des facteurs entourant la conception de la visualisation à partir des termes et des concepts propres au domaine de l'éducation vers des concepts communs aux conceptions de visualisation entre divers domaines. Au vu de l'abstraction des données et des tâches selon divers critères, on peut facilement discerner les conséquences des choix de conception individuels une à la fois. La visualisation qui en résulte est optimisée pour donner un aperçu de la tâche pour laquelle elle a été conçue à partir des données disponibles. Le guide est conçu pour être indépendant de la plate-forme et ne précise donc pas comment programmer la visualisation dans l'un des nombreux outils logiciels disponibles (p. ex. Tableau, D3, ggplot2, InfoVis, VTK); il indique seulement comment procéder à la conception.

En résumé, le présent guide vise à vous aider à :

- Comprendre à quels stades les visualisations peuvent être utiles dans votre pratique d'interprétation des données.
- Peaufiner les grands domaines d'examen en en faisant des questions de recherche précises auxquelles on peut répondre avec précision et exactitude.
- Concevoir une visualisation ayant expressément pour but de présenter les données de manière à rendre les renseignements recherchés plus évidents.
- Remettre en contexte les nouveaux domaines d'enquête afin qu'ils puissent être comparés plus facilement à ceux qui disposent déjà de solutions.

Guide de conception de la visualisation

Ce processus de conception de visualisation est adapté du flux de travail décrit dans le livre *Visualization Analysis & Design* (2014) de Munzner. Les types de données et de tâches qui se trouvent dans le domaine de l'éducation ne comprennent qu'une partie de l'objet des conseils contenus dans le livre. Le processus présenté ici est donc une version simplifiée qui peut être plus accessible aux personnes qui font des études postsecondaires (EPS). Par exemple, le présent guide porte surtout sur les étapes pertinentes aux données tabulaires, car il représente la majorité du travail effectué en éducation. Pour nos besoins, les données tabulaires sont des ensembles de données qui peuvent être organisés en lignes et en colonnes.

Le processus adapté est divisé en trois étapes générales :

1. Comprendre le problème
2. Attribuer des attributs

3. Peaufiner la visualisation

Étape 1 : Comprendre le problème

Une visualisation devrait être conçue en fonction du genre de données qu'elle sert à analyser ainsi que du genre de problème pour lequel elle fournit des renseignements. Avant de prendre des mesures pour visualiser les données, il importe de comprendre à la fois le genre de données et le genre de problème afin de simplifier la conception de la visualisation et de s'assurer que l'objectif de visualisation visé est le mieux atteint. Il y a deux volets à la compréhension du problème : connaître le type de données utilisées et définir la tâche que la visualisation vise à faciliter. Les deux sont réalisés au moyen d'un processus d'abstraction utilisant des catégories définies par Munzner. Bien qu'il y ait un léger chevauchement entre ces deux parties, il s'agit essentiellement de processus distincts qui peuvent être exécutés dans n'importe quel ordre.

Abstraction des données : Quelles données sont présentées?

La première partie de la conception d'une visualisation en réponse à une question consiste à déterminer la nature des données disponibles. Les ensembles de données disponibles ainsi que les attributs déterminent comment les données peuvent être ventilées au moyen de l'abstraction et regroupées pour répondre à la question posée. Le processus de cette abstraction commence par l'examen des données qui peuvent être disponibles auprès de la personne qui pose une question en matière d'éducation ou des autres données nécessaires à des fins pédagogiques. Chaque question nécessite un examen attentif du rôle de la personne au sein de son établissement et, par conséquent, des données dont elle dispose pour établir une visualisation. Au départ, les données nécessaires pour répondre à la question doivent être identifiées. Le présent guide suppose que les données seront déjà disponibles, mais ce processus aide à établir les données qu'il faudrait recueillir.

Tableau 1 : Résumé des catégories et des options d'abstraction des données

Abstraction des ensembles de données		Abstraction des attributs	
Catégorie d'abstraction	Options	Catégorie d'abstraction	Options
Types de données	<ul style="list-style-type: none"> • Élément • Attribut 	Sémantique de l'attribut	<ul style="list-style-type: none"> • Attribut clé • Attribut de valeur
Types d'ensemble de données	<ul style="list-style-type: none"> • Données tabulaires 	Types d'attributs	<ul style="list-style-type: none"> • Catégorique • Ordinal • Quantitatif
Disponibilité des ensembles de données	<ul style="list-style-type: none"> • Statique • Dynamique 	Orientation de l'ordre	<ul style="list-style-type: none"> • Séquentiel • Divergent • Cyclique

Le processus d'abstraction des données requiert l'identification des caractéristiques de chaque catégorie du tableau 1 pour chaque question qui nécessite une visualisation. Chaque ensemble de données devant être visualisé est classé en types de données, qui forment l'ensemble des éléments ou des attributs à visualiser, le type d'ensemble de données (dont seules les données tabulaires sont prises en compte dans le présent guide) et la disponibilité de l'ensemble de données, qui est dynamique si les données peuvent changer, ou statique, si elles ne le peuvent pas. Les **éléments** de l'ensemble de données sont l'unité d'analyse individuelle pour chaque domaine d'enquête, qui correspond à une rangée dans un ensemble de données. Les **attributs** sont les données associées à chaque élément et qui contiennent ses caractéristiques et sont divisés en attributs clés ou de valeur; ils correspondent aux colonnes d'un ensemble de données.

Par exemple, dans le cas d'un tableau qui donne le détail des étudiants et de leur rendement en matière d'évaluation dans un cours, chaque étudiant serait un élément (une unité d'analyse individuelle dans une visualisation), tandis que son rendement à chaque évaluation serait un attribut. Un tel cas se trouve dans le graphique 1, où chaque colonne est un attribut, et chaque cellule correspond à un attribut de cet étudiant. Si on se penchait plutôt sur des statistiques sommaires (moyenne, écart-type, etc.) de différentes évaluations, chaque évaluation pourrait être un élément, tandis que chaque statistique serait un attribut de cette évaluation.

Figure 1 : Exemples d'éléments et d'attributs en format tabulaire

	Attribut clé	Attributs de valeur					
	Matricule_étudiant	MPC	Heure_du_test	Citoyenneté	Date_de_naissance	Maîtrise_du_test	
Éléments	1	830744	3.99	23	Canada	1993-12-12	Compétent
	2	612032	3.06	42	Canada	1993-10-12	Sous le niveau de base
	3	691492	3.48	38	Canada	1993-02-24	Compétent
	4	548130	3.13	27	Sri Lanka	1993-09-16	Compétent
	5	708006	4.06	4	Canada	1994-06-29	Sous le niveau de base
	6	889780	0.70	19	Canada	1994-02-17	Compétent
	7	603028	3.76	50	Canada	1993-12-22	Compétent

Nota : Il s'agit d'un exemple fictif de données tabulaires montrant des éléments ainsi que des attributs clés et de valeur. Ici, chaque étudiant est un élément qui s'accompagne de six attributs connexes : le code d'étudiant, la MPC, l'heure du test, la citoyenneté, la date de naissance et la maîtrise du test. L'ID étudiant est un attribut clé, tandis que les autres sont des attributs de valeur. Les données associées à chaque étudiant sont des valeurs correspondantes pour chaque attribut.

Chaque élément d'un ensemble de données comporte des attributs qui y sont associés, lesquels sont déterminés en examinant comment les données se rapportent au monde réel dans le contexte de chaque domaine d'enquête. Plus précisément, quels attributs sont nécessaires en tant qu'attributs clés et quels

attributs sont nécessaires en tant qu'attributs de valeur? Les attributs **clés** sont ceux qui peuvent identifier un élément de façon unique, tandis que les attributs **de valeur** sont des caractéristiques de l'élément. Dans l'exemple des étudiants et de leur rendement à l'évaluation, un numéro d'étudiant serait un attribut clé, tandis que leur note dans une évaluation serait un attribut de valeur.

Les distinctions entre les attributs clés et les attributs de valeur ne sont pas intrinsèques aux données et sont souvent motivées par l'abstraction de la tâche, ce qui aide à clarifier l'objet et le but de chaque question. Les clés sont toujours reliées aux éléments de chaque visualisation. Dans le cas des questions simples, il faut généralement au moins une clé pour chaque visualisation; toutefois, plusieurs clés peuvent être nécessaires pour déterminer les attributs de regroupement qui se rattachent aux questions plus complexes. Lorsqu'on examine les regroupements d'étudiants, leur numéro d'étudiant et l'attribut utilisé pour le regroupement peuvent être des attributs clés distincts. Dans l'exemple illustré à la figure 1, la colonne « Matricule de l'étudiant » est la clé la plus évidente. Toutefois, si la visualisation regroupe également les étudiants en fonction de leur citoyenneté, il pourrait également s'agir d'une clé.

Un attribut de valeur est déterminé en établissant quelles mesures ou observations consignées sont nécessaires pour répondre à chaque question. Par exemple, pour déterminer si un groupe s'améliore dans une compétence, il faut mesurer le rendement des étudiants.

Une fois les différents types de données confirmés, l'identification du type d'ensemble de données est simple. Puisque les données tabulaires sont le principal type d'ensemble de données que l'on trouve dans l'EPS, tel sera l'objet principal de notre étude. Ce processus serait légèrement différent si des données de réseau ou des données géospatiales étaient prises en compte.

La disponibilité des ensembles de données dépend de la nature de la question. Les questions sont définies comme statiques (si les données visualisées sont toutes disponibles en même temps) ou dynamiques (si les données sont recueillies en continu et ajoutées à la visualisation). La caractérisation dynamique signifierait soit que les attributs d'un élément changeraient avec le temps, soit que des éléments seraient ajoutés ou supprimés.

Les caractéristiques de l'attribut se rangent dans l'un des trois types suivants : catégorique, ordinal ou quantitatif. Les données **catégoriques** ne suivent pas d'ordre implicite. Par conséquent, la seule opération qui peut être effectuée pour ce type de données est de déterminer si deux éléments sont identiques ou non. Si les données suivent un ordre implicite, elles peuvent être ordinales ou quantitatives. La principale différence entre ces deux données est que des données **quantitatives** peuvent faire l'objet d'opérations arithmétiques pour créer de nouvelles données, mais pas des données **ordinales**. Par exemple, une liste classée est ordinale, mais l'ajout du premier élément au troisième n'est pas un concept significatif. Le tableau 2 fournit une référence facile pour l'attribution des types d'attribut. Il est à noter que les clés ne peuvent être que catégoriques ou ordinales, et non quantitatives.

Tableau 2 : Liste des types d'attribut avec exemples

Type d'attribut	Définitions	Exemples
Catégorique	<ul style="list-style-type: none"> • Non ordonné • Ne peut être que comparé 	Corps professoral : Ingénierie, lettres et sciences humaines, droit Programme menant à un grade : Baccalauréat, baccalauréat spécialisé, maîtrise
Ordinal	<ul style="list-style-type: none"> • Ordonné • Ne peut faire partie des calculs arithmétiques 	Années d'études : Première < Deuxième < Quatrième Note à l'évaluation : A > B+ > D
Quantitatif	<ul style="list-style-type: none"> • Ordonné • Peut faire partie des calculs arithmétiques • Mesures fréquentes 	Note d'évaluation : 61 % < 80 % < 95 % Présence en classe : 558 sur 620 (90 %), 12 sur 16 (75 %)

Dans la figure 1, le matricule et la citoyenneté des étudiants sont des attributs catégoriques. Remarquez que même si le matricule étudiant est un numéro, les numéros individuels n'ont aucun lien entre eux. Il n'est pas logique de dire que l'étudiant 603028 est classé plus bas que l'étudiant 830744. La maîtrise du test est un attribut ordinal puisqu'il y a un classement clair du niveau inférieur à la base à la base à la compétence. La MPC et le temps du test sont deux attributs quantitatifs, car il serait logique de soustraire ces données, d'établir une moyenne, etc. Toutefois, l'attribut de la date de naissance n'est pas aussi simple à catégoriser, car il dépend de la façon dont l'attribut est utilisé. Si les étudiants sont regroupés par année de naissance, on pourrait alors considérer qu'il s'agit d'un attribut ordinal. Cependant, si la date de naissance était plutôt utilisée dans un calcul, elle serait considérée comme un attribut quantitatif.

L'**orientation de l'ordre** s'applique aux données ordinales et quantitatives et peut être séquentielle, divergente ou cyclique. Un attribut est séquentiel lorsque le minimum et le maximum de son étendue sont clairs. La note d'une évaluation sur 10 serait séquentielle, allant d'un minimum de 0 à un maximum de 10. Les attributs divergents sont des échelles qui comportent deux séquences qui peuvent augmenter dans des directions opposées atteignant un point zéro commun. L'écart entre la note d'un étudiant et la moyenne serait un attribut divergent. Un attribut **cyclique** est un attribut dont les valeurs entourent un point de départ. Par exemple, les jours d'un mois ou les périodes d'une année scolaire sont des exemples courants. Tous les attributs ordonnés dans la figure 1 sont séquentiels.

La sémantique des données facilite l'identification du type d'attribut et de l'orientation de l'ordre. Par exemple, les résultats de l'évaluation, comme les pourcentages ou les MPC, peuvent être clairement reconnus comme étant quantitatifs et séquentiels, les semestres peuvent être reconnus comme étant ordinaux et séquentiels, et le type de programme est catégorisé sans orientation de l'ordre.

Certains des attributs requis par la visualisation ne sont pas fournis par les données brutes. Il faut plutôt les obtenir à l'aide de calculs. Des exemples courants d'attributs dérivés (y compris les différences, les moyennes, les catégories, etc.) devraient être considérés comme des attributs de valeur supplémentaires s'il y a lieu, ou moins souvent comme des attributs clés. La nature du dérivé est liée à l'abstraction de la tâche et discutée plus en détail dans la prochaine section. Ces attributs dérivés doivent également être classés en conséquence.

Abstraction de la tâche : Pourquoi est-ce illustré?

Le processus d'abstraction des tâches clarifie les intentions et les objectifs de chaque domaine d'enquête pour lequel les données sont utilisées afin d'en améliorer la compréhension. Il est plus facile de comprendre l'objectif fondamental de la visualisation en faisant passer chaque domaine d'enquête des termes propres à un domaine, comme les notes ou les étudiants, à un domaine abstrait de termes et de concepts de visualisation, comme la répartition d'éléments. Ce processus permet également de reconnaître les similitudes entre les questions et les conceptions. De plus, l'abstraction des tâches contribue à cerner les transformations de données nécessaires qui orientent le processus d'abstraction des données. Les catégories d'abstraction des tâches figurent au tableau 3.

Tableau 3 : Sommaire des niveaux d'abstraction des actions et des options correspondantes pour chaque action

Niveau d'abstraction	Intérêt
Analyse	<ul style="list-style-type: none"> • Présences : • Découvrir : exploration/discussion
Recherche	<ul style="list-style-type: none"> • Cible connue avec emplacement inconnu • Cible et emplacement connus • Cible inconnue avec emplacement connu • Cible et emplacement inconnus
Interrogation	<ul style="list-style-type: none"> • Déterminer une cible • Comparer quelques cibles • Résumer toutes les cibles

L'abstraction de la tâche commence par l'examen de la façon dont la question est posée et de l'utilité de créer une visualisation pour donner une réponse. Parmi les exemples d'utilité, mentionnons l'amélioration de la visualisation des questions courantes ou l'approfondissement d'un cours, d'un programme, d'une institution, etc. Une fois qu'une question est concise et ciblée, il est relativement simple de classifier la tâche. Toutefois, lorsque la question n'est pas ciblée, le processus d'abstraction de la tâche peut limiter l'interrogation à l'objet de la demande.

L'abstraction de la tâche repose sur deux concepts primaires : les actions et les cibles. Une **action** est ce qu'une visualisation permet à l'utilisateur de faire avec les données et est classée en trois niveaux. Une cible est un aspect des données analysées. La conception d'une visualisation a pour but de permettre à l'utilisateur d'effectuer une action sur une cible. Ces actions correspondent à ce que la visualisation permet essentiellement à l'utilisateur de faire avec les données et peuvent être classées en trois niveaux : analyse, recherche et interrogation.

Au niveau le plus élevé, l'**analyse**, une visualisation vise soit à découvrir de nouveaux renseignements, soit à présenter des renseignements connus à d'autres. La découverte est subdivisée en exploration (lorsque l'analyse est effectuée par une seule partie) et en discussion (lorsque la visualisation est utilisée comme outil pour faciliter l'analyse collective).

Les deux autres niveaux sont déterminés lorsque nous pouvons préciser la nature de la cible. La **cible** est ce qu'un utilisateur tente d'analyser à partir d'une visualisation. Si c'est l'ensemble des données qui est concerné, la cible peut être une tendance, une valeur aberrante ou une caractéristique plus précise. Dans le cas d'un attribut à valeur unique, la cible pourrait correspondre à la répartition des données ou des éléments extrêmes des données. Lorsqu'il y a plusieurs attributs, la cible peut être une dépendance ou une corrélation. Les types de cibles figurent au tableau 4.

Tableau 4 : Résumé des types de cibles et de l'état des tâches auxquels ils peuvent s'appliquer

Type de cible	État
<ul style="list-style-type: none"> • Tendance • Caractéristique 	Peut s'appliquer à toutes les données
<ul style="list-style-type: none"> • Valeur aberrante • Extrême 	Il peut s'agir d'un seul élément ou de quelques éléments
<ul style="list-style-type: none"> • Répartition 	Les données ont un seul attribut de valeur
<ul style="list-style-type: none"> • Dépendance • Corrélation 	Les données ont plusieurs attributs de valeur

Une fois le type de cible identifié, les actions peuvent être classées selon le prochain niveau, soit ses classifications de recherche. Le niveau de **recherche** exige de déterminer si l'identité de la cible est connue avant d'examiner la visualisation et si l'emplacement de la cible dans les données est connu. L'emplacement dépend souvent de la spécificité de la question. Par exemple, les questions axées sur un sous-ensemble de données en particulier ont des emplacements connus, tandis que les questions génériques ont tendance à s'accompagner d'emplacements inconnus.

Le niveau le plus bas, l'**interrogation**, dépend en partie du nombre de cibles. L'action pourrait consister à déterminer une seule cible, à comparer plusieurs cibles ou à résumer chacune d'entre elles. Par exemple, si vous tentiez de voir s'il y a une augmentation ou une diminution des taux de diplomation au fil du temps, vous détermineriez (interrogation sur les actions) une tendance (cible). Par ailleurs, si vous tentiez d'établir la performance d'un groupe d'étudiants par rapport à l'ensemble de la classe, vous compareriez (interrogation) les répartitions (cibles).

Comme nous l'avons mentionné précédemment, cette méthode d'abstraction des tâches présente l'avantage supplémentaire d'évaluer la spécificité de la tâche actuelle. Si une visualisation est créée pour une question qui ne correspond pas parfaitement à ces catégories d'abstraction des tâches, la question doit probablement être précisée. Une question comme « Existe-t-il des cliques au sein d'une équipe étudiante? » n'a pas pu être soigneusement abstraite, car une « clique » ne constituerait pas une cible de caractéristique facilement identifiable à partir d'un ensemble de données. La question pourrait être peaufinée : « Quelles équipes d'étudiants manifestent des perceptions incohérentes? » La nouvelle question est plus spécifique et il est plus facile d'y répondre à l'aide d'une visualisation.

Les tâches et l'abstraction des données se chevauchent lorsque l'on tient compte des données dérivées. Ces données sont calculées en plus des données brutes recueillies. Parmi les exemples de données dérivées, mentionnons les moyennes des notes d'une évaluation ou d'un étudiant à l'autre, les dénombrements des champs des intervalles et les différences entre les notes ou les évaluations des étudiants. Le chevauchement entre les deux abstractions se produit lorsque les données dérivées sont ramenées de nombreux attributs

de valeur à quelques-uns, comme dans le calcul de la moyenne des résultats d'une évaluation; les nombreuses notes des évaluations sont alors réduites à une valeur moyenne unique. Il importe de préciser les cibles qui sont définies ou comparées afin de ne pas en réduire le nombre. Par exemple, si la cible d'une visualisation était un étudiant, il serait préférable d'éviter toute réduction du nombre des élèves. On pourrait par exemple calculer la moyenne d'une évaluation ou compartimenter les étudiants dans un histogramme. Dans le cas des données tabulaires, comme dans le cas d'un tableau où chaque étudiant est une ligne et chaque colonne est une mesure d'évaluation, on peut concevoir de façon utile la réduction en se demandant si les lignes (étudiants) ou les colonnes (évaluations) sont celles qui sont réduites.

Dans l'ensemble, les étapes principales à suivre pour un type de données et une tâche sont les suivantes :

Abstraction des données

1. Déterminer les éléments et attributs des données.
2. Définir les clés et valeurs associées aux attributs.
3. Déterminer le type d'attribut et l'orientation de l'ordre (le cas échéant) pour chaque attribut.
4. Déterminer quelles données doivent être obtenues pour la tâche.

Abstraction de la tâche

1. Déterminer si le but de la visualisation est de présenter, d'explorer ou de créer une discussion.
2. Déterminer la cible.
3. Déterminer si vous cherchez une cible précise et si son emplacement est connu.
4. Établir de façon finale si vous tentez d'identifier une cible, si vous comparez quelques cibles ou si vous résumez toutes les cibles.

Étape 2 : Attribuer des attributs

Une fois le processus d'abstraction des données terminé, il devrait en résulter une liste d'attributs, chacun étant une clé ou une valeur d'un type particulier (de type catégorique, ordinal ou quantitatif). De plus, selon l'abstraction de la tâche, il pourrait y avoir un ordre de préséance des attributs. Par exemple, en examinant les moyennes et les écarts-types d'un ensemble d'évaluations, la tâche pourrait consister soit à comparer les moyennes, soit à répartir les résultats. Si l'on examine les moyennes, la moyenne comporterait un précédent plus élevé, alors que si la diffusion des résultats était l'objectif principal, cet attribut aurait un précédent plus élevé. L'autre attribut vaut peut-être encore la peine d'être visualisé, mais de façon secondaire.

Un tel tableau peut faire en sorte qu'il soit possible d'établir de façon assez mécanique une première version lors d'une visualisation au moyen de marques visuelles et de voies. Une **marque** est un élément géométrique de base qui représente un élément comme des points sur des diagrammes de dispersion ou

des lignes dans des diagrammes à barres. Les **voies** sont des façons différentes de contrôler l'apparence d'une marque, comme la position, la forme et la couleur. Les voies sont utilisées avec des marques pour représenter l'ampleur (données quantitatives) ou l'identité (données catégorisées ou ordinales). Certaines voies ont été démontrées empiriquement pour être lues de façon plus ou moins exacte. Le tableau 5 énumère les voies aux fins de l'ampleur et de l'identité du plus lisible au moins lisible.

Tableau 5 : Liste ordonnée des voies (adaptée de la figure 5.1 dans Munzner, 2014)

	Ampleur	Identité
Plus précis	Position sur une échelle commune (horizontale ou verticale)	Position (région spatiale)
	Position sur une échelle non alignée	Teinte de la couleur
	Longueur	Forme
	Inclinaison/angle	
	Superficie (2D)	
	Luminance de couleur	
	Saturation des couleurs	
	Volume (3D)	
Moins précis		

Remarquez que la position se trouve au début des deux listes du tableau 5, ce qui indique dans quelle mesure la position est mieux placée que bon nombre des autres voies et qu'il faut y accorder une attention et une priorité spéciales. La position peut constituer une façon exacte de représenter les données quantitatives et de distinguer efficacement les catégories en séparant les notes par régions. Un exemple commun de ces deux types utilisés se trouve dans un diagramme à barres catégorisées, dans lequel la position sur une échelle commune sert à représenter la valeur des quantités de chaque barre tout en regroupant celles-ci, ce qui montre qu'elles font partie du même groupe.

Certaines voies peuvent également se nuire et devraient être traitées avec soin. Par exemple, si une voie de position horizontale est utilisée sur les marques de concert avec une voie de position verticale, cela pourrait ressembler à une voie régionale qui aurait des conséquences imprévues sur la perception des données.

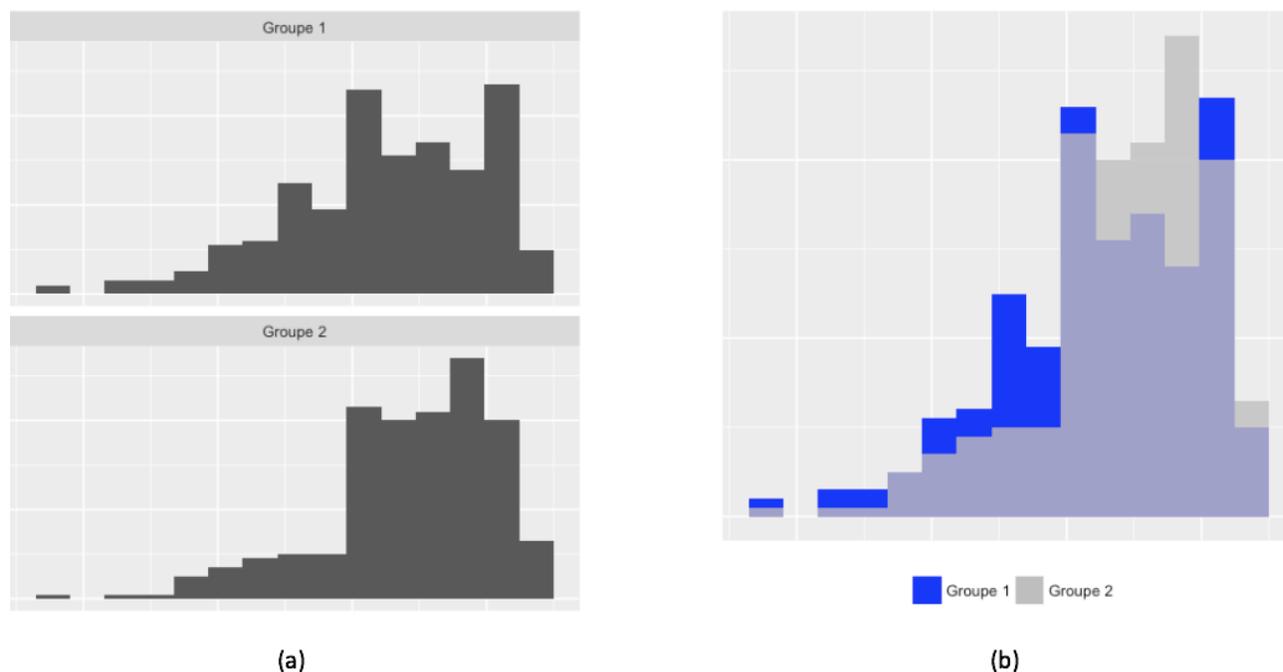
Il faut aussi accorder une attention particulière à la couleur. La voie de la couleur peut être séparée selon la teinte, la luminance et la saturation. La teinte d'une couleur (bleu, rouge, jaune, etc.) peut être utilisée avec beaucoup d'effet pour distinguer les catégories, mais elle est médiocre pour représenter l'information quantitative. La luminance et la saturation d'une couleur (son caractère foncé et sa luminosité) sont des façons utiles de transmettre de l'information quantitative, mais elles doivent être utilisées avec prudence compte tenu de leur faible classement au tableau 5. Bien que la couleur puisse être bien utilisée à des fins

de comparaison entre les marques, elle ne permet pas à l'utilisateur de lire avec exactitude de l'information quantitative. De plus, lorsque vous choisissez des couleurs, assurez-vous que votre palette de couleurs fonctionne pour les personnes aveugles. [ColourBrewer](#) est une excellente ressource en ligne pour les palettes de couleurs. En outre, voici une bonne règle de base : une visualisation peut toujours fonctionner aux fins prévues lorsqu'elle est vue en gris (sans teintes). Ainsi, les trois aspects de la couleur sont suffisamment distincts et accessibles, quel que soit le support dans lequel la visualisation est présentée.

Les voies peuvent être assignées à des attributs de valeur en déterminant si elles sont quantitatives ou catégoriques et en parcourant graduellement la liste de priorité. Toutefois, les choix devraient permettre de distinguer facilement les clés dans la mesure requise par la tâche. En général, si vous avez une clé, votre visualisation ressemblera à une liste. Un diagramme à barres peut être conçu comme une liste de barres, chacune représentant une valeur quantitative. S'il y a deux clés, la visualisation ressemblera probablement à une matrice. Une carte thermique est l'exemple le plus courant : l'on y trouve des lignes et des colonnes pour identifier les clés, et la valeur de chaque cellule est représentée par une voie quantitative, souvent de la luminosité. Avec trois clés, vous pouvez utiliser une visualisation 3D, mais il n'est pas recommandé de le faire. On pourrait également choisir de réutiliser certaines voies, comme la région spatiale, en optant toutefois pour une orientation différente, tout en constatant des effets secondaires supplémentaires. Plus de trois clés nécessitent des techniques plus avancées, comme la subdivision récursive, où de petites sous-visualisations sont subdivisées dans la visualisation globale. Par exemple, un histogramme ou un diagramme à secteurs dans les barres d'un diagramme à barres.

Un autre aspect à considérer lors de l'attribution d'attributs de valeur consiste à déterminer si la visualisation doit être à facettes ou superposée. L'établissement de facettes affiche des visualisations similaires les unes les autres séparées par une clé. Lorsqu'il y a plusieurs clés, c'est une bonne façon d'utiliser plusieurs fois une visualisation à une clé ou à deux clés. La figure 2a illustre un exemple de visualisation avec une clé utilisée à plusieurs reprises pour une autre clé avec des valeurs de groupe 1 et de groupe 2. Ce peut être particulièrement utile pour comparer un attribut de valeur spécifique sur une clé. La superposition est une autre approche utile, surtout lorsque la tâche clé consiste à comparer deux cibles. La superposition affiche une visualisation au-dessus d'une autre en utilisant la même échelle et le même espace, souvent en réduisant la transparence de la visualisation du haut. La figure 2b illustre un exemple où les valeurs d'une autre clé sont superposées à des fins de comparaison.

Figure 2 : Exemples de comparaison de groupes semblables dans un ensemble de données



Nota : En a), les ensembles de données sont en facettes, tandis que b) illustre comment les ensembles de données peuvent être superposés.

Si vous utilisez l'établissement de facettes ou la superposition, ayez recours à un ordre logique pour ce groupement. Il peut y avoir un ordre implicite, mais si ce n'est pas le cas, un ordre peut être imposé. Par exemple, lorsque l'on compare des groupes d'étudiants non ordinaux, le nombre d'étudiants dans chaque groupe peut constituer un ordre possible.

Dans l'ensemble, les principales étapes à suivre pour attribuer les attributs sont les suivantes :

1. Déterminer si les attributs sont des clés ou des valeurs et si les attributs de valeur sont quantitatifs, ordinaux ou catégoriques.
2. Attribuer les voies en ordre en fonction des attributs de valeur qui sont les plus importants pour la visualisation, tout en tenant compte de l'interférence entre les voies.
 - Décider si la visualisation nécessite une facette ou une superposition.

Étape 3 : Peaufiner la visualisation

Une fois qu'une visualisation a été conçue, l'étape finale consiste à l'évaluer et à l'améliorer. Une fois que les voies sont attribuées aux données respectives, la visualisation qui en résulte peut ou non ressembler à une visualisation familière. À ce stade, il importe de vérifier si la visualisation créée convient bien à la tâche dont elle voulait donner un aperçu. Les visualisations courantes sont prolifiques, car elles ont été utilisées dans de nombreuses tâches, mais elles ne sont pas toujours la meilleure visualisation pour une fin donnée.

Il arrive fréquemment que lorsqu'une visualisation conçue pour une tâche particulière est élaborée et affichée, l'information qui se présente le plus facilement ne convienne peut-être pas à la tâche qu'elle visait à éclairer. Si tel est le cas, il existe deux solutions possibles. La première consiste à revoir l'étape 2 et à trouver d'autres façons de représenter les données qui présentent plus fidèlement et plus facilement l'information que vous cherchez tout en tenant compte de ce qui, comme on le sait, ne fonctionne pas. Il s'agit d'une méthode comportant des possibilités d'essais et d'erreurs.

L'autre solution consiste à adapter la visualisation à une tâche différente et à revenir à l'étape 1. Cette visualisation peut donner un aperçu d'une autre question qu'il vaut également la peine de poser. L'avantage de cette deuxième option est que la visualisation a déjà été créée. L'établissement d'une visualisation est un processus très itératif et il est courant de refaire les étapes précédentes lorsque vous voyez une visualisation finale des données.

Si la visualisation fournit le bon type d'information, assurez-vous de fournir tous les détails nécessaires pour permettre à l'utilisateur d'interpréter rapidement la visualisation. Ces renseignements comprennent les légendes, les étiquettes, les titres, les marques de graduation sur les axes, le texte annoté sur les marques et tout autre détail qui accélérerait l'interprétation de la visualisation (voir les figures 3 et 5 dans les exemples). Le formatage de la visualisation, comme la couleur de fond, les lignes de la grille d'axe et d'autres aspects, peut avoir une incidence importante sur la lisibilité de l'information. Il convient parfois d'utiliser des voies plus d'une fois dans la même visualisation. Par exemple, un diagramme à barres de différentes couleurs peut déjà avoir une légende, mais il pourrait être utile d'ajouter du texte à chaque barre pour faciliter l'interprétation. Cette étape peut souvent consister à mettre à l'essai différentes options et à voir ce qui fonctionne le mieux, ou même à donner aux consommateurs éventuels des options quant à la façon dont ils utilisent les données.

Limites du guide et pratiques exemplaires

Comme nous l'avons mentionné précédemment, le présent guide porte surtout sur les données tabulaires. D'autres types de données, comme les données de réseau ou les données géographiques, nécessiteraient des approches légèrement modifiées. L'outil utilisé pour créer la visualisation pourrait également restreindre celle-ci dans le cas des données tabulaires. Bon nombre des outils logiciels spécialisés pour la visualisation peuvent traiter de très grands ensembles de données (p. ex., l'examen de données pour des

institutions entières). Toutefois, les visualisations de cette taille pourraient ne pas fonctionner aussi bien en format réel ou interactif. Il se peut que les outils de visualisation plus utilisés, comme Microsoft Excel, prennent plus difficilement en charge des visualisations plus complexes ou des données qui changent fréquemment. C'est une raison majeure pour laquelle le présent guide a été établi sans que ses auteurs aient un outil en particulier à l'esprit, puisque ce processus de conception d'une visualisation s'applique aux ensembles de données de toute taille. Cependant, selon les outils et les compétences disponibles, seules certaines conceptions de visualisation peuvent être accessibles.

En outre, les visualisations fonctionnent le mieux lorsqu'il s'agit de fournir de l'information qui peut aider les utilisateurs à prendre des décisions fondées sur des données. Ces décisions ne peuvent être prises uniquement au moyen de la visualisation. Il y a une étape d'interprétation entre le recours à la visualisation et la prise de décision, au cours de laquelle l'utilisateur applique le point de vue de la visualisation à ses besoins. Par exemple, lorsqu'un directeur de programme examine les cours les plus fréquemment choisis parmi ses étudiants, il est maintenant mieux informé de la nécessité de modifier le programme. L'information tirée de la visualisation pourrait même laisser croire qu'aucun changement n'est nécessaire.

Par conséquent, lorsque vous tentez de trouver en quoi les visualisations peuvent faciliter votre pratique, ne cherchez pas les défis que les visualisations peuvent résoudre. Concentrez-vous plutôt sur le genre d'information qui peut le mieux vous aider à relever ces défis. Une question comme « Le programme enseigne-t-il et évalue-t-il les résultats que nous croyons qu'il enseigne et évalue? » vaut la peine que l'on y réponde, mais pourrait plutôt être reformulée comme suit : « Où enseigne-t-on et évalue-t-on les résultats attendus? » Une visualisation conçue en fonction de la dernière question peut ensuite aider à répondre à la première par une réflexion supplémentaire du corps professoral et du personnel concerné.

Exemples du Guide

Exemple 1 : « Les groupes sous-représentés présentent-ils le même taux de diplomation que les groupes fortement représentés? »

Étape 1 : Abstraction des données et des tâches

La tâche consiste à comparer les tendances des taux de diplomation des étudiants. La cible correspondrait alors aux tendances et l'action serait de comparer. Dans ce cas, l'emplacement est connu puisque les données pertinentes seraient connues au préalable, mais la cible serait inconnue puisqu'il existe plusieurs groupes qu'il pourrait être utile de comparer. Le tableau de données abstraites serait le tableau 6. Il convient de souligner que nous avons choisi l'année d'obtention du diplôme comme valeur ordinale, car c'était une clé. Le temps est traité comme une valeur discrète dans ce cas-ci, compte tenu du contexte dans lequel une classe de finissants aurait peu d'effet sur les années successives, bien que le temps puisse être considéré comme quantitatif pour une tâche différente.

Tableau 6 : Tableau de données abstraites pour une question sur les taux de diplomation

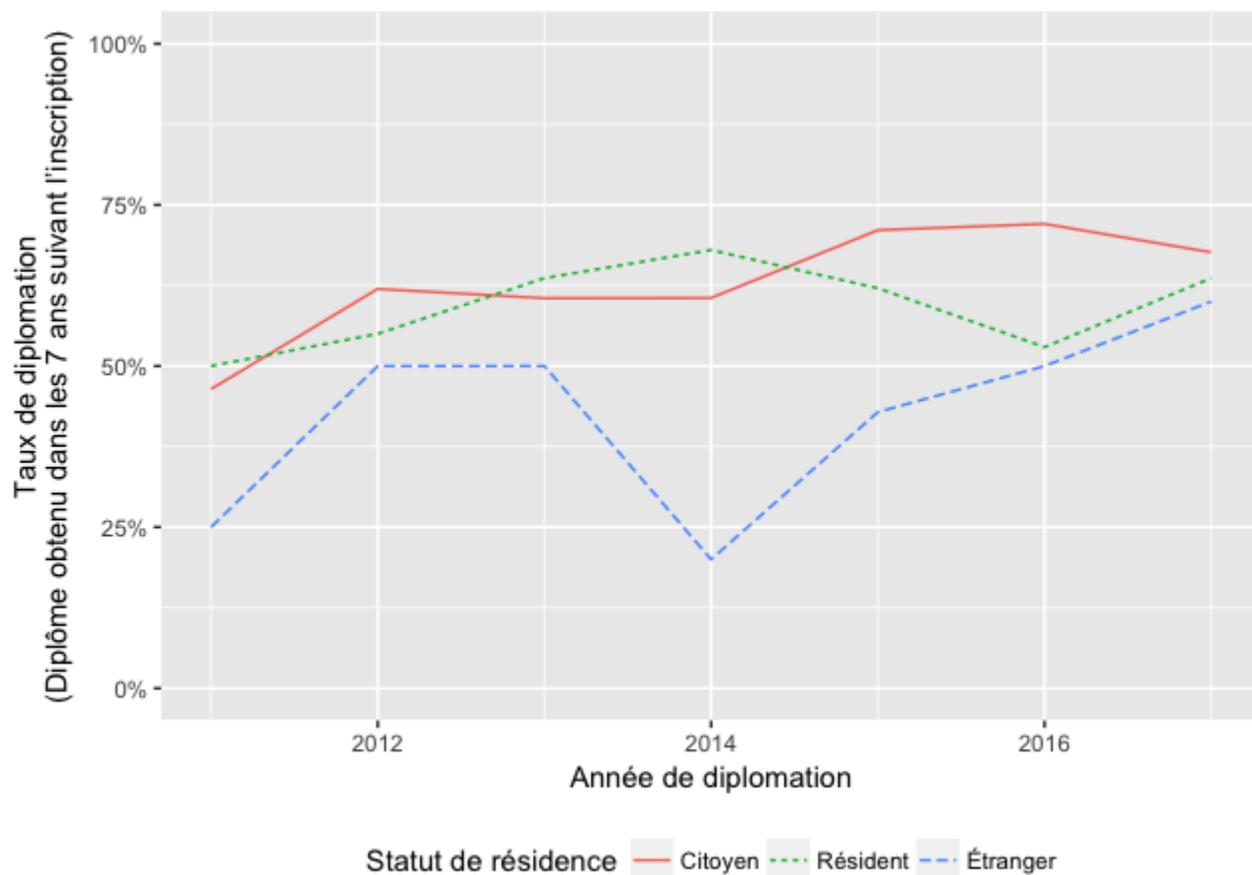
Attribut	Sémantique de l'attribut	Type d'attribut
Groupe démographique	Clé	Catégorique
Année d'obtention du diplôme	Clé	Ordinal, séquentiel
Taux de diplomation (dérivé)	Valeur	Quantitatif, séquentiel

Étape 2 : Attribuer des attributs

Les voies peuvent ensuite être attribuées en consultant le tableau 4. Il existe deux clés : le groupe démographique et l'année d'obtention du diplôme. La région spatiale, la couleur ou la forme peuvent leur être attribuées, car ce sont des attributs d'identité. Puisque la tâche consiste à comparer les tendances au fil des ans, il faut consulter les différents groupes démographiques de la même région pour faciliter la comparaison. Par conséquent, la couleur est utilisée pour la région du groupe démographique qui s'applique à l'année d'obtention du diplôme. Il existe un attribut de valeur, de sorte que la position peut être utilisée sur une échelle commune. Il en résulte des marques de couleur positionnées dans un espace 2D selon l'année et le taux de diplomation.

En ce qui concerne la marque visuelle, vous pouvez choisir un point ou une ligne pour représenter les données. Ce choix refléterait la différence entre un diagramme de dispersion et un diagramme à barres, respectivement. L'utilisation de points permet de relier les notes à des lignes pour souligner que les notes du même groupe sont liées. En ce qui concerne les deux voies de position, il n'y a pas de raison implicite dans les abstractions des données qui indique si le taux de diplomation ou l'année de diplomation doit occuper les voies horizontales ou verticales. Toutefois, par convention, le temps a été choisi pour occuper la voie horizontale. Sinon, la visualisation pourrait être inutilement obtuse pour les utilisateurs qui sont habitués à une certaine norme de visualisation. La figure 3 montre la visualisation possible de ces choix.

Figure 3: Visualisation de « Les groupes sous-représentés présentent-ils le même taux de diplomation que les groupes fortement représentés? »



Étape 3 : Peaufiner la visualisation

Comme cette question traite expressément des groupes sous-représentés, il pourrait également être utile d'inclure un attribut quantitatif supplémentaire indiquant le pourcentage de la population que représente ce groupe. Comme on peut s'y attendre, la proportion varierait d'une année à l'autre. Une autre voie peut être attribuée à cet attribut si elle ne nuit pas à la voie verticale occupée par le taux de diplomation. La marque actuelle pourrait être augmentée ou d'autres marques peuvent être ajoutées. Une façon d'augmenter la marque actuelle serait d'utiliser la voie de profil de sorte que chaque point soit une tranche d'un diagramme à secteurs. Il serait également possible d'utiliser la voie des positions horizontales, où la marque de chaque année est placée horizontalement, en proportion de chaque région. Un tout nouvel ensemble de marques pourrait par ailleurs être utilisé lorsqu'un diagramme à barres pour chaque année-région représente la proportion d'étudiants. Les options sont nombreuses, mais le processus offre une façon

structurée de mettre à l'essai des approches différentes, qui pourraient être viables. La meilleure façon de trouver la visualisation la plus efficace consiste à tenter d'appliquer différentes options pour établir ce qui fonctionne et ce qui ne fonctionne pas.

Exemple 2 : « Quels sont les cours les plus fréquemment choisis pour les étudiants d'un groupe démographique en particulier? »

Étape 1 : Abstraction des données et des tâches

Dans ce cas, la tâche consiste à déterminer les cours qui constituent des extrêmes ou des valeurs aberrantes dans les taux de choix de cours. La cible réside ici dans les valeurs aberrantes et l'action consiste à déterminer. Dans le cas qui nous occupe, la cible et l'emplacement sont inconnus. L'ensemble de données abstraites serait présenté au tableau 7. Le taux de choix de cours correspond ici au pourcentage du groupe démographique ayant suivi le cours. Dans ce cas, nous disposons également de certains renseignements sur la répartition des cours en fonction de l'année d'études.

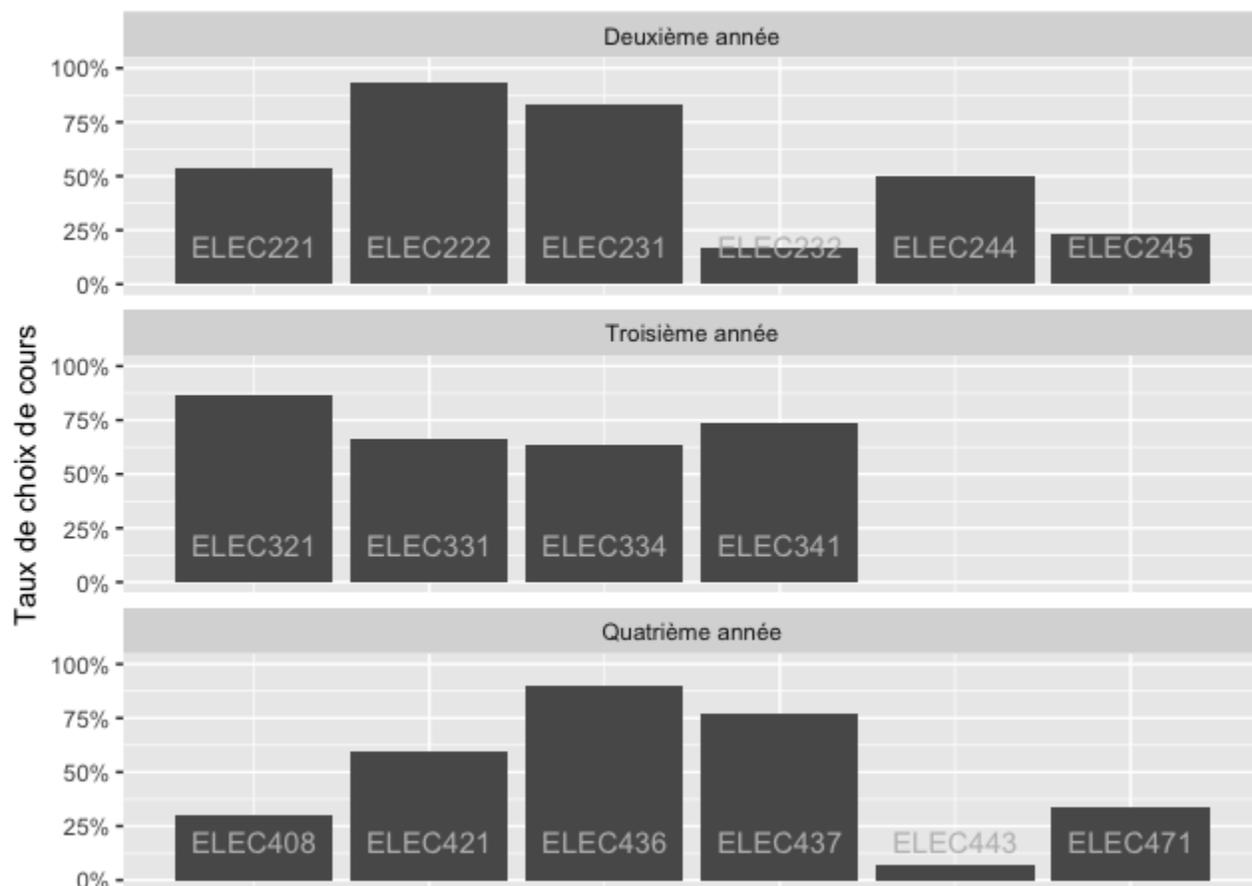
Tableau 7 : Tableau de données abstraites pour la question sur le choix de cours

Attribut	Sémantique de l'attribut	Type d'attribut
Taux de sélection des cours du groupe démographique (%)	Valeur	Quantitatif, séquentiel
Année d'étude	Clé	Ordinal

Étape 2 : Attribuer des attributs

La façon la plus simple de procéder consisterait à dresser une liste des taux de sélection, regroupés par année d'études. La voie de position est utilisée pour l'attribut de valeur, tandis que les barres sont divisées en régions pour l'année d'études, classées par ordre de la deuxième à la quatrième année. Cette version initiale de la visualisation se trouve à la figure 4.

Figure 4: Première version de la visualisation pour « Quels sont les cours les plus fréquemment choisis pour les étudiants d'un groupe démographique en particulier? »



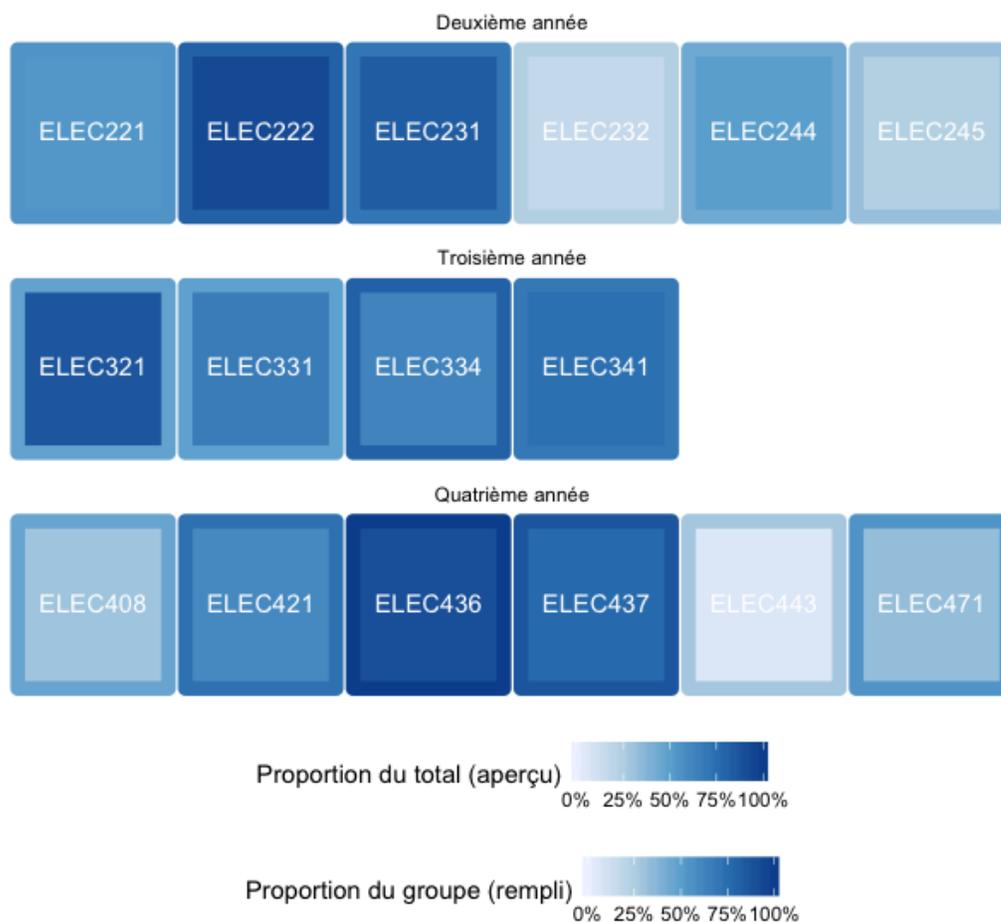
Étape 3 : Peaufiner la visualisation

Après avoir créé la version initiale de la visualisation, comme le montre la figure 4, deux problèmes sont apparents. Tout d'abord, si l'utilisateur peut dire quels cours sont les plus ou les moins populaires parmi ce groupe démographique, il n'a aucun moyen de savoir si la situation est différente de l'ensemble du groupe. Par conséquent, il devrait peut-être exister un moyen d'établir une comparaison, d'une manière moins importante que les taux de sélection du groupe démographique. Ensuite, bien que les taux de sélection soient faciles à comparer à ceux des autres cours pour chaque année d'études, il est moins évident de déterminer quels sont les cours les plus et les moins choisis de toutes les années.

Il existe un certain nombre de façons de réattribuer des voies pour régler ces problèmes. Nous pourrions ajouter des barres distinctes pour chaque cours afin de comparer les taux de sélection du groupe

démographique. Cette façon de procéder pourrait toutefois détourner l'attention de l'objectif principal de la visualisation. Une autre approche consisterait à utiliser la voie de la couleur pour les taux de sélection totaux, mais il serait alors plus difficile de comparer les taux de sélection du groupe à l'ensemble des taux, car l'un se trouverait sur la voie de la position alors que l'autre serait dans celle de la couleur. Nous pouvons plutôt utiliser la couleur pour les deux voies tout en faisant en sorte que l'une des deux ressorte un peu moins, comme le montre la figure 4. Ici, chaque barre est remplie de la couleur correspondant à la proportion du groupe démographique qui a suivi le cours, tandis que celle-ci est illustrée par la proportion de la population totale qui a suivi le cours selon la légende ci-dessous. Dans cette version, nous pouvons tout de même déterminer clairement quels cours sont suivis le plus ou le moins souvent, et nous pouvons ensuite établir une comparaison pour voir si cette tendance en matière de sélection de cours est anormale par rapport à la proportion totale. Par exemple, environ la moitié des étudiants ont pris le cours ELEC 221, qu'ils appartiennent ou non au groupe démographique. Peu d'étudiants ont pris ELEC 232, sans égard au groupe démographique. Toutefois, nous pouvons également constater que le cours ELEC 408 est moins populaire dans ce groupe démographique que dans l'ensemble de la population. En revanche, le cours ELEC 321 est choisi par une proportion beaucoup plus importante du groupe démographique que dans l'ensemble de la population. Il pourrait donc valoir la peine d'approfondir la question.

Figure 5: Version mise à jour de la visualisation pour « Quels sont les cours les plus fréquemment choisis pour les étudiants d'un groupe démographique particulier? »



Mentionnons qu'il n'est pas aussi facile de comparer les valeurs exactes des taux de sélection du groupe démographique. Cela suffit toutefois, puisque la comparaison n'est pas le but de la visualisation. L'objectif principal consiste à déterminer ceux qui sont le plus et le moins choisis, ce que cette méthode permet d'établir de façon manifeste. Toutefois, si cette information est requise, elle pourrait être ajoutée sous forme d'éléments de texte plus petits sur chaque bloc.

D'autres éléments pourraient également être pris en compte. Si certains cours de l'ensemble sont des préalables l'un à l'autre, il se pourrait qu'ils soient triés en régions pour rendre ces liens facilement visibles.

Références

- Bertin, J. (1981). *La graphique et le traitement graphique de l'information*. Berlin : Walter de Gruyter.
- Börner, K. et Polley, D. E. (2014). *Visual insights: A practical guide to making sense of data*. Boston : MIT Press.
- Munzner, T. (2014). *Visualization Analysis and Design*. CRC Press.
- Sivanand, A. et Frank, B. (2017). Information Visualisation in Education: A Review of Current Tools and Practices. *Proceedings of the Canadian Engineering Educational Association*.
- Sivanand, A. (2017). Supporting post-secondary educational data usage in the assessment process with information visualization. Thèse de doctorat, Université Queen's.
- Tufte, E. R. et Graves-Morris, P. R. (1983). The visual display of quantitative information, vol. 2, n° 9. Cheshire, CT : Graphics Press.

Glossaire

Action : Une action est ce qu'une visualisation permet à l'utilisateur de faire avec les données. Les actions sont classées en trois niveaux : analyse, recherche et interrogation.

Analyse : L'un des trois niveaux d'action, l'analyse, est le niveau auquel une visualisation est censée présenter de l'information connue à autrui ou découvrir de nouvelles informations. La découverte de nouveaux renseignements est subdivisée en *exploration* (lorsque l'analyse est effectuée par une seule partie) et en *discussion* (lorsque la visualisation est utilisée comme outil pour faciliter l'analyse de groupe).

Attribut : Données associées à chaque élément et contenant ses caractéristiques. Les attributs sont divisés en attributs clés et attributs de valeur; ils correspondent aux colonnes d'un ensemble de données. Les attributs clés peuvent être divisés en catégoriques ou ordinaux. Les attributs de valeur peuvent être catégoriques, ordinaux ou quantitatifs.

Attribut clé : L'un des deux types d'attributs, un attribut clé est un attribut qui permet d'identifier un élément de façon unique. Les clés sont toujours reliées aux éléments de chaque visualisation. Au moins une clé est habituellement requise, mais plusieurs clés peuvent être nécessaires pour identifier les attributs de regroupement pour les questions plus complexes. Les clés ne peuvent être que catégoriques ou ordinales, et *non* quantitatives.

Attribut de valeur : Il s'agit de l'un des deux types d'attribut. Un attribut de valeur est une caractéristique d'un élément. Les valeurs sont déterminées en établissant quelles mesures ou observations enregistrées sont nécessaires pour répondre à une question. Les attributs de valeur peuvent être catégoriques, ordinaux ou quantitatifs.

Catégorique : L'un des trois sous-types d'attributs ainsi que les données ordinales et quantitatives; les données catégoriques ne comportent pas d'ordre implicite.

Cible : Une cible est un aspect des données analysées ou de ce qu'un utilisateur tente d'analyser à partir d'une visualisation. Si ce sont les données dans leur ensemble qui sont concernées, la cible peut être une tendance, une valeur aberrante ou une caractéristique plus précise. Dans le cas d'un attribut à valeur unique, la cible pourrait être la distribution des données ou des éléments extrêmes des données. Lorsqu'il y a plusieurs attributs, la cible peut être une dépendance ou une corrélation.

Cyclique : L'une des trois catégories d'orientation de l'ordre, un attribut est cyclique si ses valeurs forment une boucle autour du point de départ.

Divergent : L'une des trois catégories d'orientation de l'ordre, un attribut est divergent lorsqu'il comporte deux séquences qui peuvent augmenter en sens opposés et atteindre un point zéro commun.

Élément : Un élément est l'unité d'analyse individuelle pour chaque domaine d'enquête. Les éléments correspondent aux lignes d'un ensemble de données.

Interrogation : Le niveau le plus bas de définition d'une action dans le processus d'abstraction des tâches consiste à déterminer une cible, à comparer plusieurs cibles ou à résumer toutes les cibles, ce qui dépend du nombre de cibles.

Marque : Une marque est un élément géométrique de base qui représente un élément comme des points sur des diagrammes de dispersion ou des lignes dans des diagrammes à barres. L'apparence d'une marque peut être contrôlée par des voies.

Ordinal : L'un des trois sous-types d'attributs, de même que les données catégoriques et quantitatives; les données ordinales sont des données implicitement mises en ordre qui ne *peuvent pas* faire l'objet de calculs arithmétiques pour créer de nouvelles données.

Orientation de l'ordre : S'applique aux données ordinales et quantitatives et peut être séquentielle, divergente ou cyclique.

Quantitatif : L'un des trois sous-types de données, avec les données catégoriques et ordinales; ces données sont des données implicitement ordonnées qui *peuvent* faire l'objet de calculs arithmétiques pour créer de nouvelles données.

Recherche : L'un des trois niveaux de définition d'une action dans le processus d'abstraction des tâches. Au niveau de la recherche, il s'agit de déterminer si l'identité de la cible est connue avant d'examiner la visualisation, et si l'emplacement de la cible parmi les données est connu. L'emplacement dépend souvent de la spécificité de la question.

Séquentiel : Une des trois catégories d'orientation de l'ordre, un attribut est séquentiel lorsqu'il s'agit d'une plage dont le minimum et le maximum sont clairs.

Valeur : Le point de référence associé à un attribut d'un élément donné.

Voie : Les voies sont l'une des façons de contrôler l'apparence d'une marque, comme la position, la forme et la couleur. Les voies sont utilisées avec des marques pour représenter l'ampleur (données quantitatives) ou l'identité (données catégoriques ou ordinales).

Résumé des catégories et des options d'abstraction des données

Abstraction des ensembles de données		Abstraction des attributs	
Catégorie d'abstraction	Options	Catégorie d'abstraction	Options
Types de données	<ul style="list-style-type: none"> • Élément • Attribut 	Sémantique de l'attribut	<ul style="list-style-type: none"> • Attribut clé • Attribut de valeur
Types d'ensemble de données	<ul style="list-style-type: none"> • Données tabulaires 	Types d'attributs	<ul style="list-style-type: none"> • Catégorique • Ordinal • Quantitatif
Disponibilité des ensembles de données	<ul style="list-style-type: none"> • Statique • Dynamique 	Orientation de l'ordre	<ul style="list-style-type: none"> • Séquentiel • Divergent • Cyclique

