

L'analyse des données à l'appui de la prise de décision

Formateur : Patrick Boily

Diapositives : P. Boily (IACS, DAL, uOttawa), J. Schellinck (Sysabee, DAL, AI Guides), J. Stroud (AI Guides)
S. Davies (DAVHILL, DAL), B. Conway-Smith (Sysabee), M. Kashef

Plan de l'exposé

Module 1

Processus de prise de décision

Module 2

Raisonnement, éléments de preuve, renseignements, données

Module 3

Les personnes, l'éthique des données, et la loi

Module 4

L'intelligence économique et l'analytique

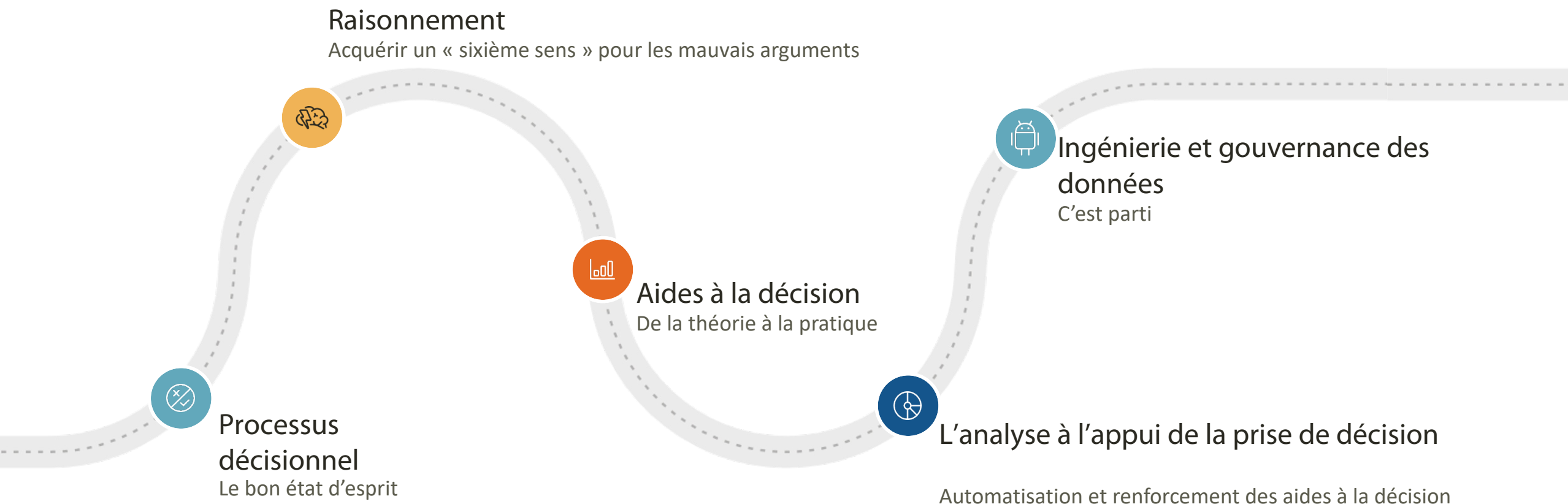
Module 5

L'analyse et l'appui décisionnel

Module 6

L'ingénierie et la gouvernance des données

Parcours de formation



Le contexte : les personnes, l'éthique des données et la loi

Décisions prises au cours de l'année écoulée

Faut-il rénover le toit cette année ou l'année prochaine?

Faut-il aller en Nouvelle-Écosse en pleine pandémie?

Devrais-je acheter une voiture?

Mon fils doit-il aller en internat?

Table ronde



Brève introduction

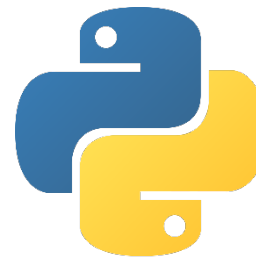
Expérience

Pourquoi ce cours?

Quelles décisions avez-vous prises/prenez-vous?

Exercice : Prendre une décision

Outils d'analyse et de données



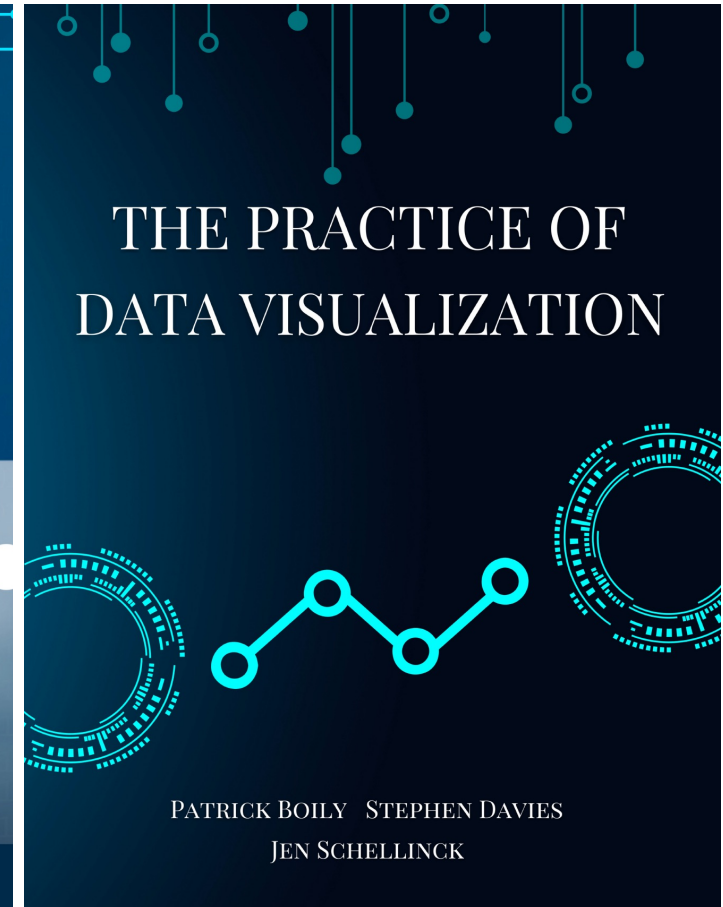
Références recommandées

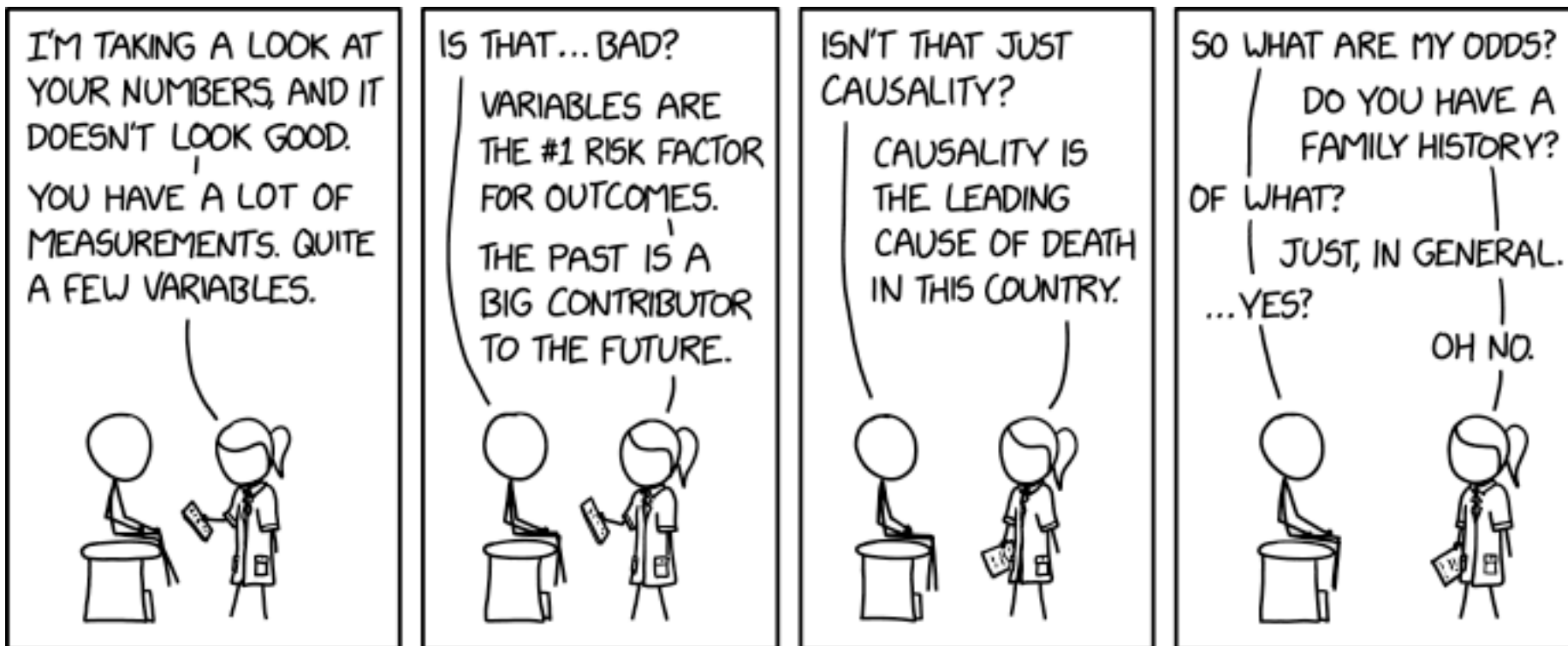
Data Understanding, Data Analysis, and Data Science

idlewyldanalytics.com

The Practice of Data Visualization

pracicedataviz.com





Faites un don pour nous aider à trouver un remède à la causalité.

Personne ne devrait avoir à souffrir d'événements à cause d'autres événements. [xkcd n° 2602]



Module 1

Processus décisionnel

Choisir, décider, agir

Lorsque nous prenons une décision, nous choisissons d'agir d'une manière ou d'une autre (ou de plusieurs autres).

Nous espérons que l'action choisie contribuera à la réalisation de nos objectifs.

La vie tourne autour de deux choses

La chance

(qui échappe à votre contrôle)



Les décisions

(que vous contrôlez)



Jouer le tout pour le tout

Situation

- Il reste 26 secondes au Super Bowl, l'équipe adverse mène par 4 points, le ballon est à la ligne de 1 verge.
- « Tout le monde » pense que vous allez remettre le ballon à votre demi-offensif vedette.

Options

A : Courir avec le ballon (1 jeu). **Risque** : Ne pas marquer avant la fin du temps réglementaire.

B : Lancer le ballon puis courir s'il le faut (2 jeux). **Risque** : 2 % de risque d'interception.



The worst play call in NFL history will continue to haunt Seahawks in 2015

The worst play call in the history of the NFL will continue to haunt the Seattle Seahawks during the 2015 NFL season.

DON BANKS • JUL 20, 2015

L'« erreur » d'Hillary Clinton

Lisez l'article de [Vanity Fair](#) :

« Tous les électeurs qui ont coûté l'élection à Clinton pourraient tenir dans un stade de football de taille moyenne »

Était-ce de la malchance, ou une erreur?



Acheter les droits de Star Wars

United Artists, Universal et Disney ont dit non merci.

20th Century Fox a dit oui.

Des recettes de 775 millions de dollars pour le premier film et de 10 milliards pour la série complète.

20th Century Fox : de l'intelligence ou de la chance?



Résultat

		Qualité des résultats	
		Bons	Mauvais
Qualité de la décision	Bonne	Récompense bien méritée	Malchance
	Mauvaise	Pur hasard	Ils ont couru après

Plusieurs futurs, un seul passé

Quand on envisage le futur : on voit un arbre à plusieurs branches

Quand on regarde le passé : on passe toutes les branches à la tronçonneuse





Décrire le processus décisionnel

Avoir confiance en son instinct

Liste des avantages et des inconvénients

En pilotage automatique

Consensus du comité

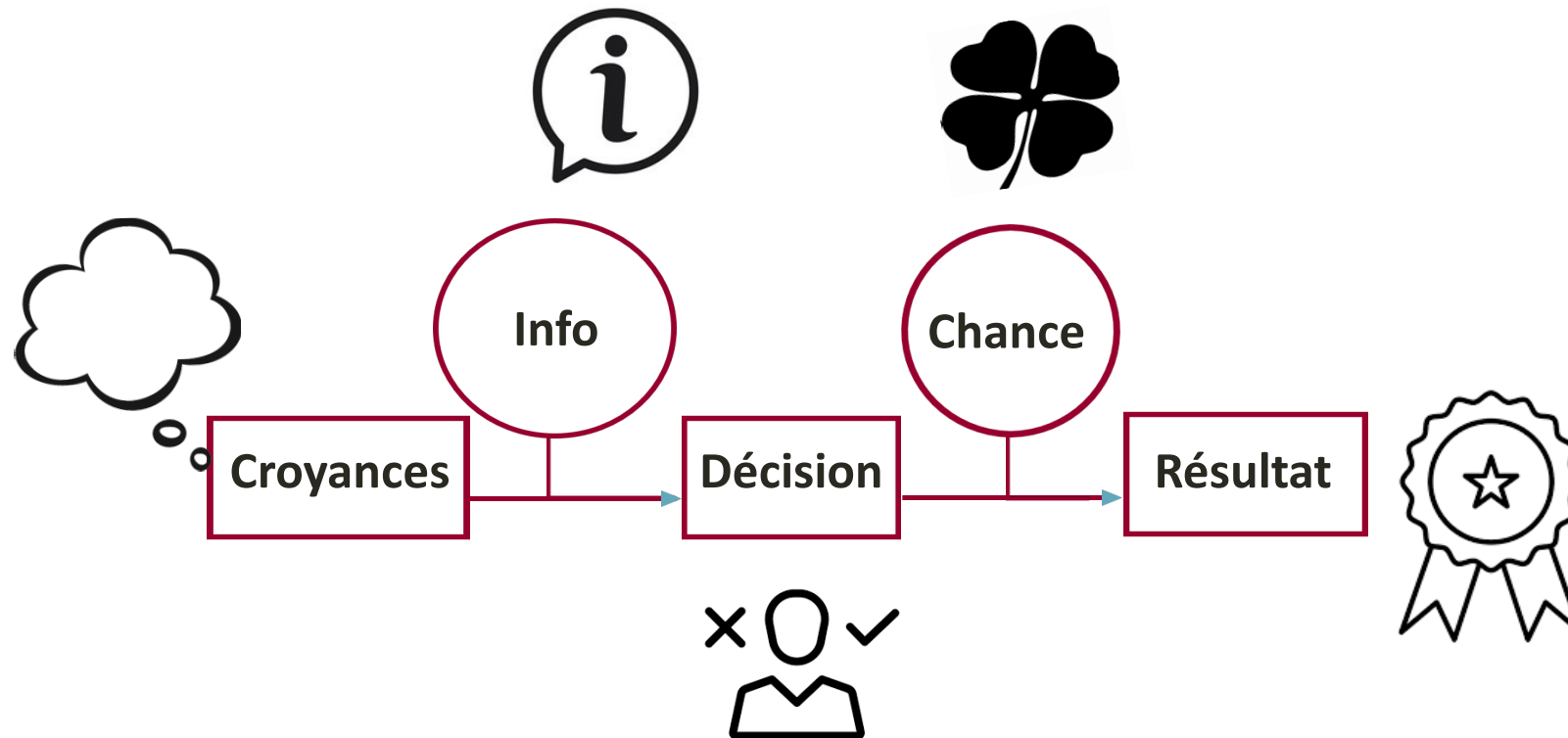
Laisser faire le hasard

Meilleure démarche (« boule de cristal »)

Objectifs :

- Possibilité de l'appliquer à plusieurs reprises, de la même manière à chaque fois.
- Possibilité de l'enseigner à autrui.
- Possibilité de l'améliorer au fil du temps (dans certains cas).
- Possibilité d'examiner rétrospectivement si on s'en est bien servi.
- Automatisation.

Chance et information



Types de préjugés

Biais variable

Biais du survivant

L'effet IKEA

Biais d'ancrage

Biais de confirmation

Sophisme de la planification

Biais du progrès

Biais de disponibilité

L'effet IKEA

Accorder trop de crédit au travail qu'on fait soi-même.



Solution : évitez d'imposer une solution **simplement parce** que vous en êtes l'auteur.

Biais variable

Biais du
survivant

L'effet IKEA

Biais d'ancrage

Biais de
confirmation

Sophisme de la
planification

Biais du progrès

Biais de
disponibilité

Biais d'ancrage

Également connu sous le nom de « biais de la première impression » – conclusions hâtives



Solution : prenez du recul et recalibrez votre réflexion.

Biais variable

Biais du
survivant

L'effet IKEA

Biais d'ancrage

Biais de
confirmation

Sophisme de la
planification

Biais du progrès

Biais de
disponibilité

Biais de disponibilité

Accorder trop d'importance à la première solution réaliste.



Solution : ne vous attachez pas à la première solution **seulement** pour gagner du temps.

Biais variable

Biais du
survivant

L'effet IKEA

Biais d'ancrage

Biais de
confirmation

Sophisme de la
planification

Biais du progrès

Biais de
disponibilité

Systemes 1 et 2 : Aveugles à notre aveuglement

Systeme 1 : Automatique

- Fonctionnement généralement bon, mais beaucoup de biais systémiques.
- Systeme qui favorise les « conclusions hâtives ».
- Nous sommes aveugles à notre aveuglement sur l'usage du systeme 1.

Systeme 2 : Délibéré

- Nécessite de l'attention.
- L'expression « porter attention » convient bien, car il s'agit d'un acte délibéré.
- L'application de règles logiques exige des efforts.



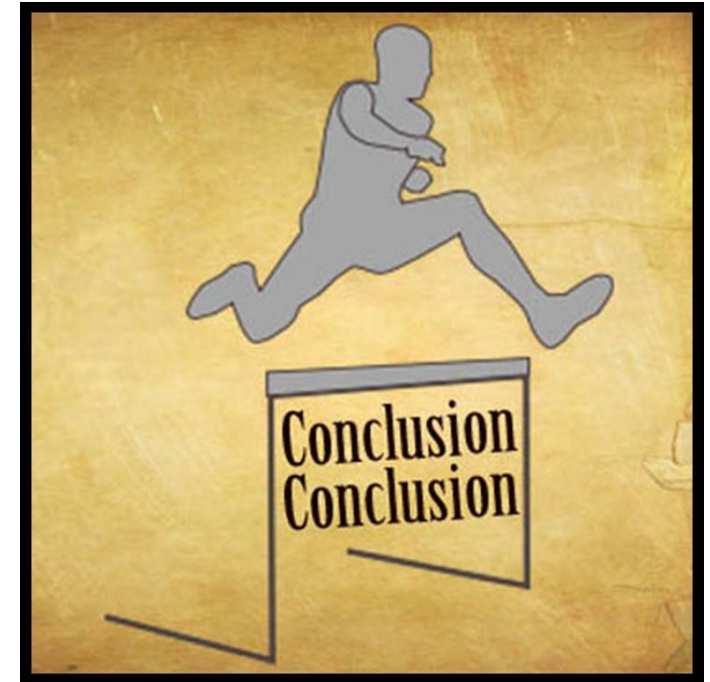
Ce qu'on voit, c'est tout ce qu'il y a à voir

Il n'y a rien d'autre à voir

- Nous (et nos patrons) sautons aux conclusions en nous appuyant sur peu d'indices matériels.

L'effet Dunning-Kruger

- « Vous ne savez pas ce que vous ne savez pas »
- Le fait d'en savoir peu sur quelque chose nous pousse à avoir trop confiance en nos connaissances et nos capacités.



Exercice : Décision sur un traitement médical

Une nouvelle maladie se propage dans votre collectivité. La bonne nouvelle est que des traitements sont mis au point.

Cependant, l'efficacité des traitements varie. L'hôpital a choisi de créer deux groupes de travail (équipe 1 et équipe 2) pour étudier les traitements et formuler une recommandation.

Étudiez les données probantes qui existent sur les traitements et formulez une recommandation.

Instinct grégaire

« Individuellement on est intelligent, **mais en foule on est stupide, on panique comme une horde d'animaux, et tu le sais.** »



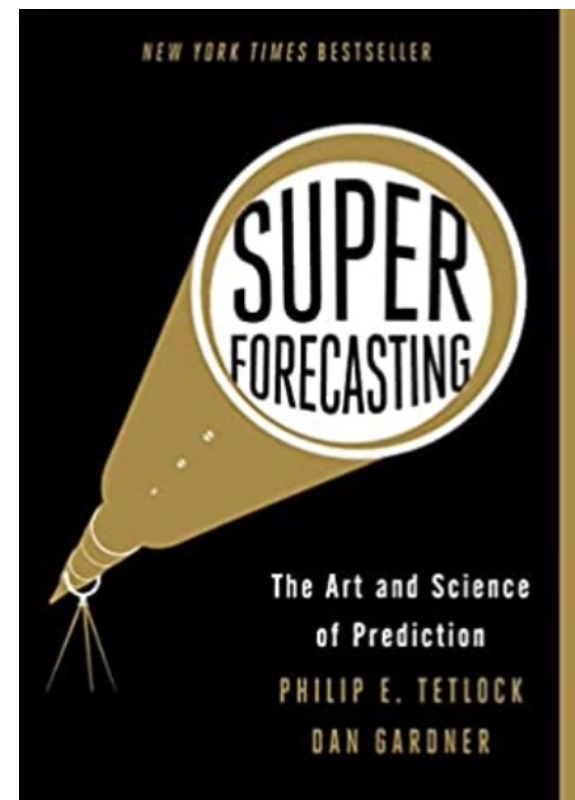
La sagesse collective



Matrice de l'évaluation des risques

LIKELIHOOD	Almost Certain					
	Likely					1
	Possible		4	1		
	Unlikely		7	2	1	
	Rare		2	3	1	
		Insignificant	Minor	Moderate	Major	Catastrophic
		IMPACT				

Tournoi de l'IARPA



Profil du super-prévisionniste

Aucune qualification ou expérience particulière.

Ce qui compte, ce n'est pas QUI ils sont.

C'est surtout COMMENT ils s'y prennent :

- curieux
- autocritiques
- humbles
- disposés à apprendre

Soyez un renard (et non un hérisson)



Vieux dicton grec :
« Il sait bien des tours, le renard. Le hérisson n'en connaît qu'un, mais c'est tout un truc. »

Intelligence amplifiée

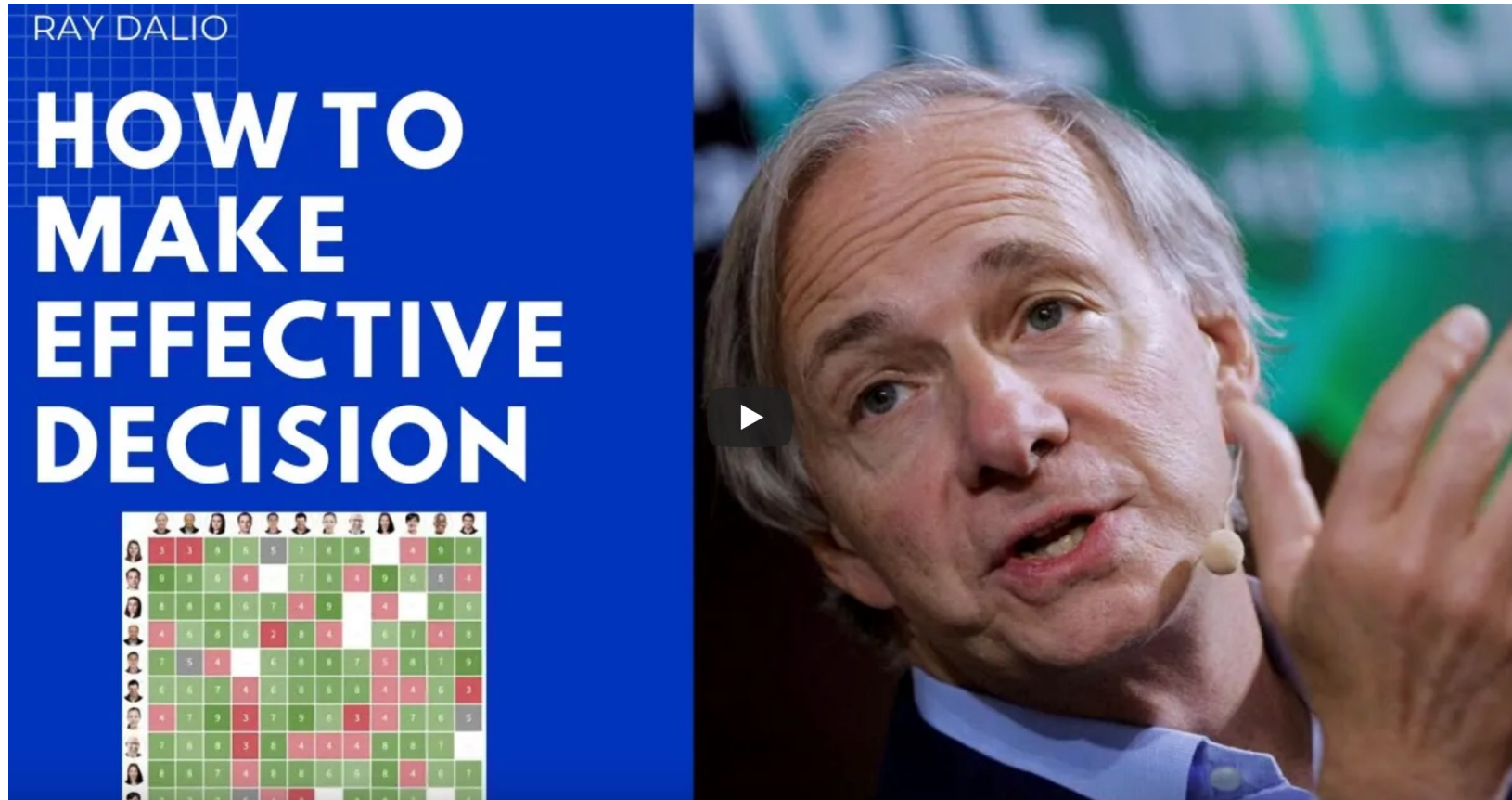
Intelligence amplifiée = super-prévisionnistes + algorithmes

On ne donne pas le même poids à tous les prévisionnistes :

1. Donnez plus de poids aux meilleures prévisions.
2. Faites passer les prévisions à l'extrême (c'est-à-dire rapprochez-les de 0 ou 100 % de probabilité).

Par exemple, si la prévision est de 70 %, faites-la passer à 85 %; si elle est de 30 %, à 15 %.

Ray Dalio : « le rassembleur de points » et l'IA



Éviter les caisses de résonance au travail

1. Demandez aux gens leur avis.
2. Ne donnez pas votre avis quand vous demandez celui d'autrui.
3. Demandez des avis de manière indépendante (surtout avant que le patron ne s'exprime).

Conseils pour prendre rapidement des décisions personnelles

1. Test de bonheur

La décision aura-t-elle un grand effet sur votre bonheur dans 1 an?

2. Le pire des scénarios

« Quel est le pire qui puisse arriver? »

3. Proximité

L'année prochaine, préférez-vous des vacances à Rome ou à Paris? Y a-t-il un mauvais choix?

4. Pas d'autre choix

« Si c'était la seule option, serais-je satisfaite? »

Cadres pour la prise de décision

Prise de décision solide

Parcours de planification adaptative dynamique

Planification par objectifs et résultats clés

Score « RICE »



Module 2

Raisonnement, éléments de preuve, renseignements, données

Comment être un renard (et non un hérisson)

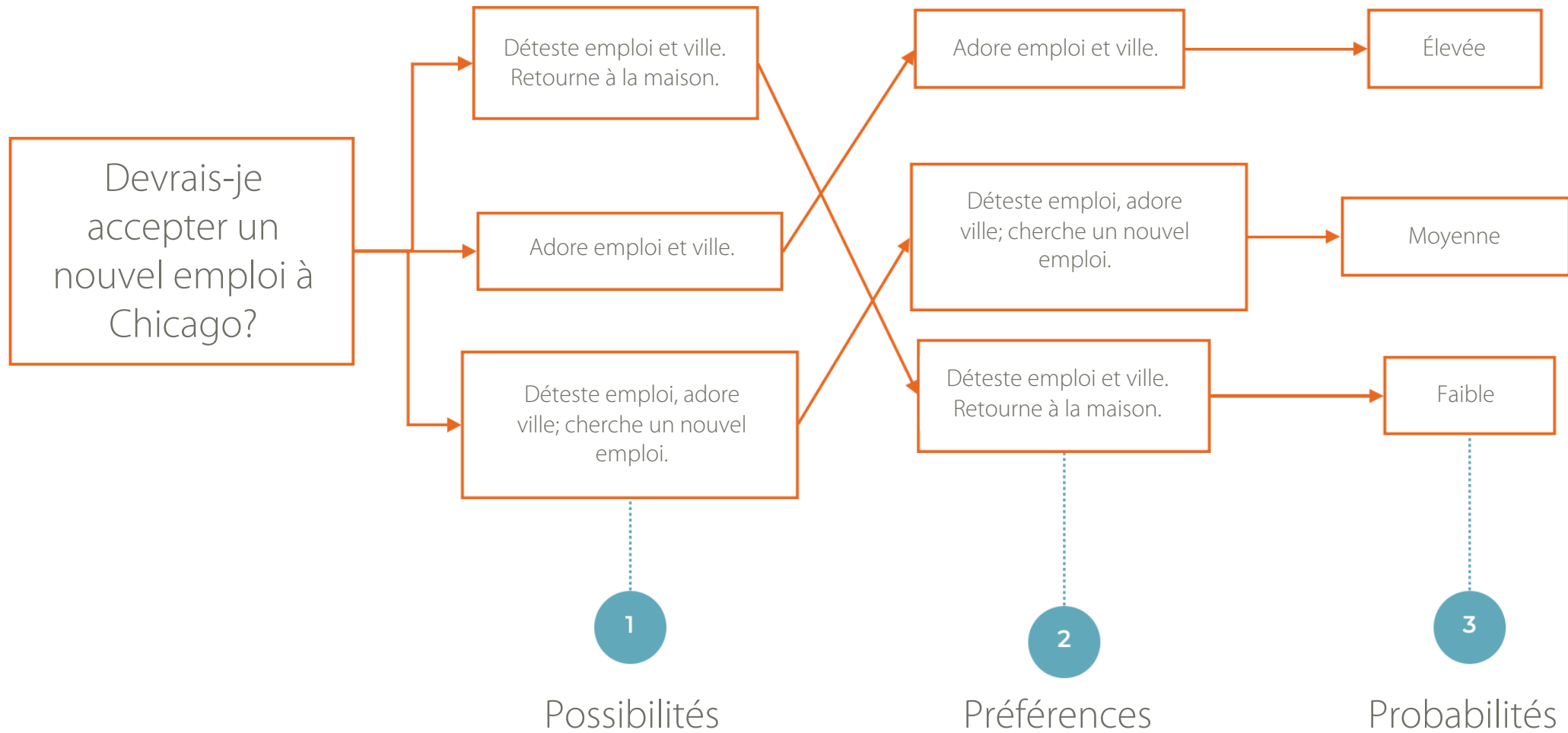
Nous avons déjà dit que, en tant que décideur, vous devez être un renard. De facto, cela signifie que l'analyse est votre travail.

Comment pouvez-vous expliquer aux autres pourquoi l'analyse – et par conséquent votre travail – est importante dans la prise de décision?

Nous allons entrer dans les détails.



Le guide des décisions d'Annie Duke



Un certain effort est nécessaire

Le système 1 est facile.

Le système 2 est difficile.

Cela en vaut-il la peine? Oui. Le savoir est un bien précieux.



**À quoi ressemble l'analyse dans
votre organisation?**

En quoi consiste l'analyse au GdC?

Tirer des conclusions?

Rassembler et présenter des éléments de preuve (tableaux croisés dynamiques)?

Proposer des options?

Formuler des opinions, des hypothèses, des croyances, des recommandations?

Faire avancer ses propres objectifs?

Activités d'analyse typiques

L'analyse porte nécessairement **sur** quelque chose. Nous analysons la **situation** ou le **problème** :

- Rassembler les faits et les éléments de preuve.
- Résumer les faits.
- Examiner et évaluer les faits.
- Regrouper les faits.
- Générer de nouveaux énoncés ou de nouvelles hypothèses.
- Décomposer les concepts en concepts plus simples.
- Construire des concepts complexes en partant de concepts simples.
- Définir les concepts.
- Dériver de nouveaux faits par le raisonnement.
- Déterminer si les énoncés sont vrais (faits) ou faux.
- Déterminer le degré de confiance qu'on a dans la véracité ou la fausseté d'un énoncé.

Thème commun : les faits!

Analyse de salon



Pensée critique

Nous voulons vous montrer en quoi la **pensée critique** (soutenue par l'analyse, le raisonnement, la déduction) est importante.

Pourquoi est-il important d'être rigoureux et méthodique?

Ceci n'est pas un cours de logique, MAIS...

En fin de compte, on vise par le raisonnement à atteindre la (une?) **vérité**, c'est-à-dire avoir suffisamment de faits à portée de main pour ne pas prendre de mauvaises décisions.



Exercice : Raisonnement et arguments

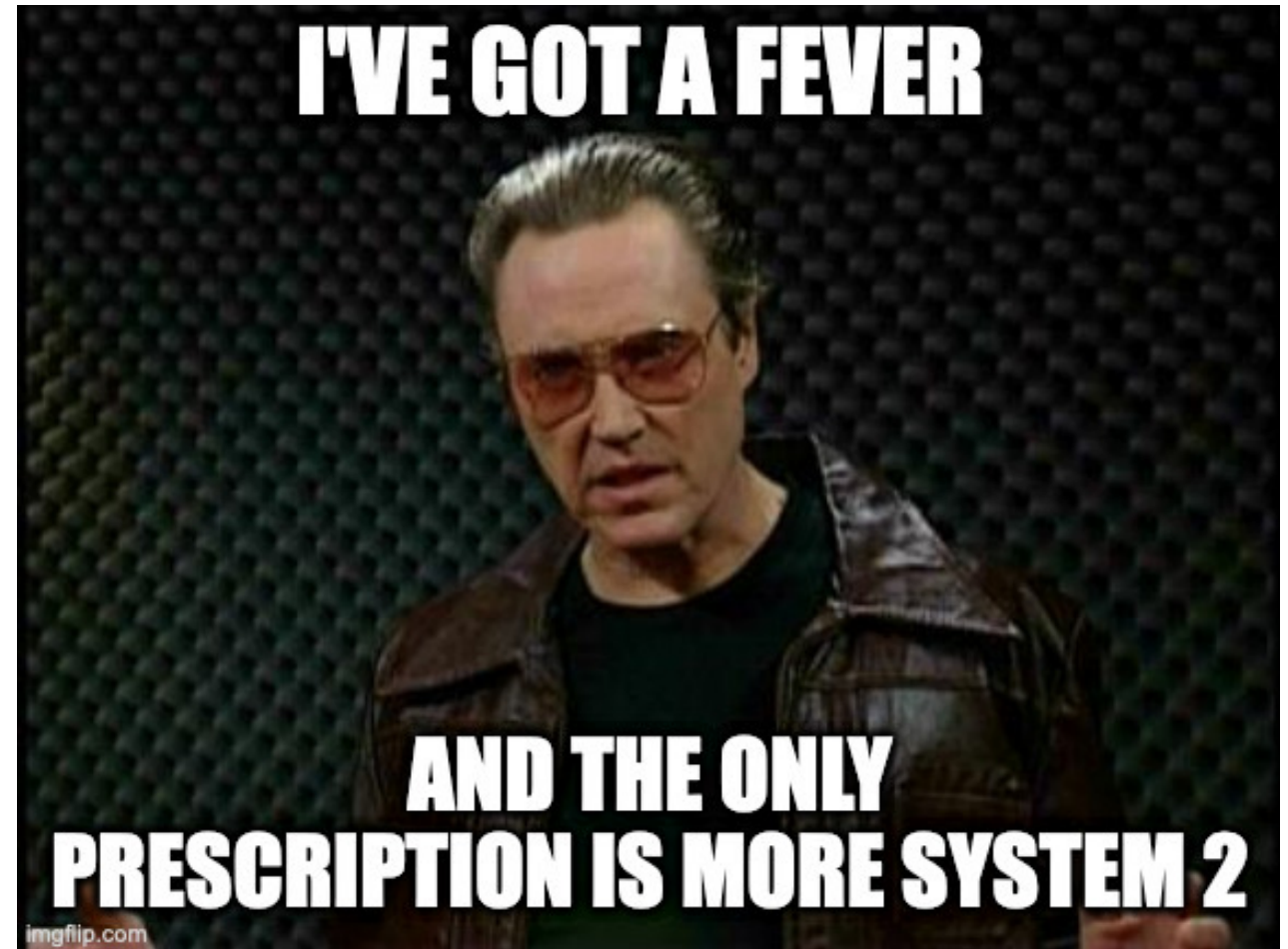
Ces énoncés tiennent-ils la route? Si ce n'est pas le cas, quels sont leurs défauts? Comment pourriez-vous les améliorer?

- a) Le vaccin contre la COVID multiplie les hospitalisations puisque la moitié des personnes hospitalisées étaient vaccinées.
- b) La mise en service du grand collisionneur de hadrons était une erreur, car soit il détruit la Terre, soit il ne la détruit pas; c.-à-d. une probabilité de 50 %, ce qui est beaucoup trop risqué.
- c) Nous savons que la Terre n'est pas plate, car aucune des autres planètes que nous connaissons n'est plate.
- d) Vous ne devriez pas voter aux prochaines élections, car une seule voix ne peut rien changer.
- e) Pour réduire la congestion sur les routes, il faut réduire le nombre de voies, car s'il y a moins de voies, les gens chercheront d'autres modes de transport.
- f) Les mesures de sécurité prises dans les aéroports sont proportionnelles au risque, car il est normal d'attendre quelques heures si cela signifie que mon avion ne sera pas détourné.

Dénominateur commun des arguments boiteux

Les principaux dénominateurs communs qui ressortent de cet exercice :

1. Pas assez de données vraies ou précises – il faut plus de faits!
2. Il faut appliquer encore plus le système 2.



Le raisonnement en vaut-il la peine?

Plus de rigueur demande **plus d'efforts** (le système 2 demande plus de travail que le système 1).

Mais il y a un prix à payer si on n'applique PAS les techniques d'analyse :

- On est incapable de distinguer ce qui est vrai de ce qui ne l'est pas.
- On se trompe, ce qui entraîne du gaspillage, etc.

Si nos convictions ne correspondent pas à la réalité, on prend de mauvaises décisions.

Options de systèmes (revues et corrigées)

Systeme 1 : Décisions automatiques

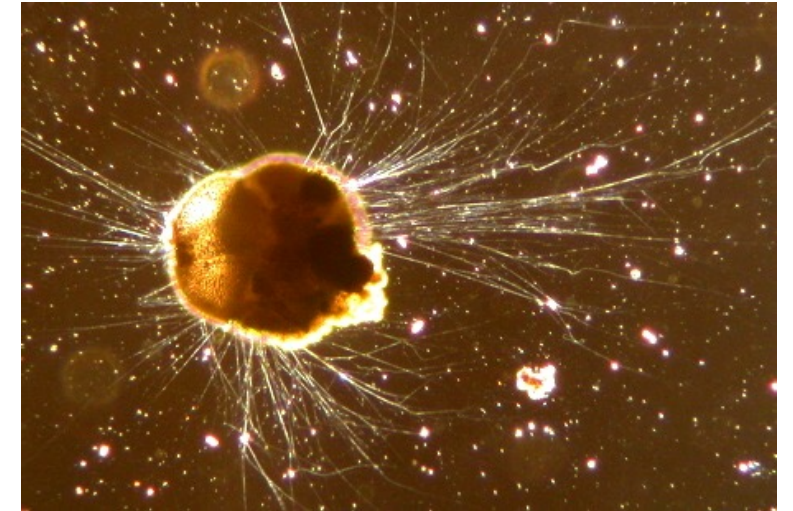
Systeme 2 : Méthode scientifique [environnement contrôlé]

Mais aussi...

Systeme 2 : Tout le reste (c'est-à-dire votre travail)

Nous devons utiliser tous les types de raisonnement...
en mettant l'accent sur ce qui est **plausiblement** vrai.

La science : se focaliser sur la généralisation



La science : contexte expérimental particulier

Les techniques d'analyse des données scientifiques ne sont parfois pertinentes que :

- dans un **contexte expérimental très particulier**;
- pour **certains types de données**.

Maintenant que les données sont beaucoup plus répandues et utilisables, nous devons **développer et adapter** ces techniques.

Nous devons peut-être sortir de la « mentalité scientifique » (**sans sacrifier la rigueur pour autant**).

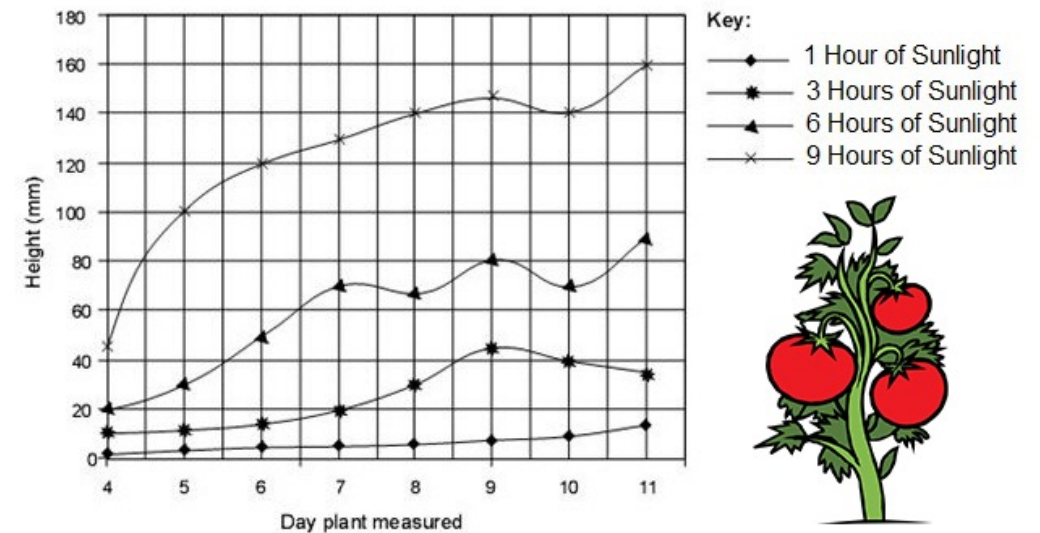


La science : types de variables

Dans un cadre expérimental, nous travaillons généralement avec des :

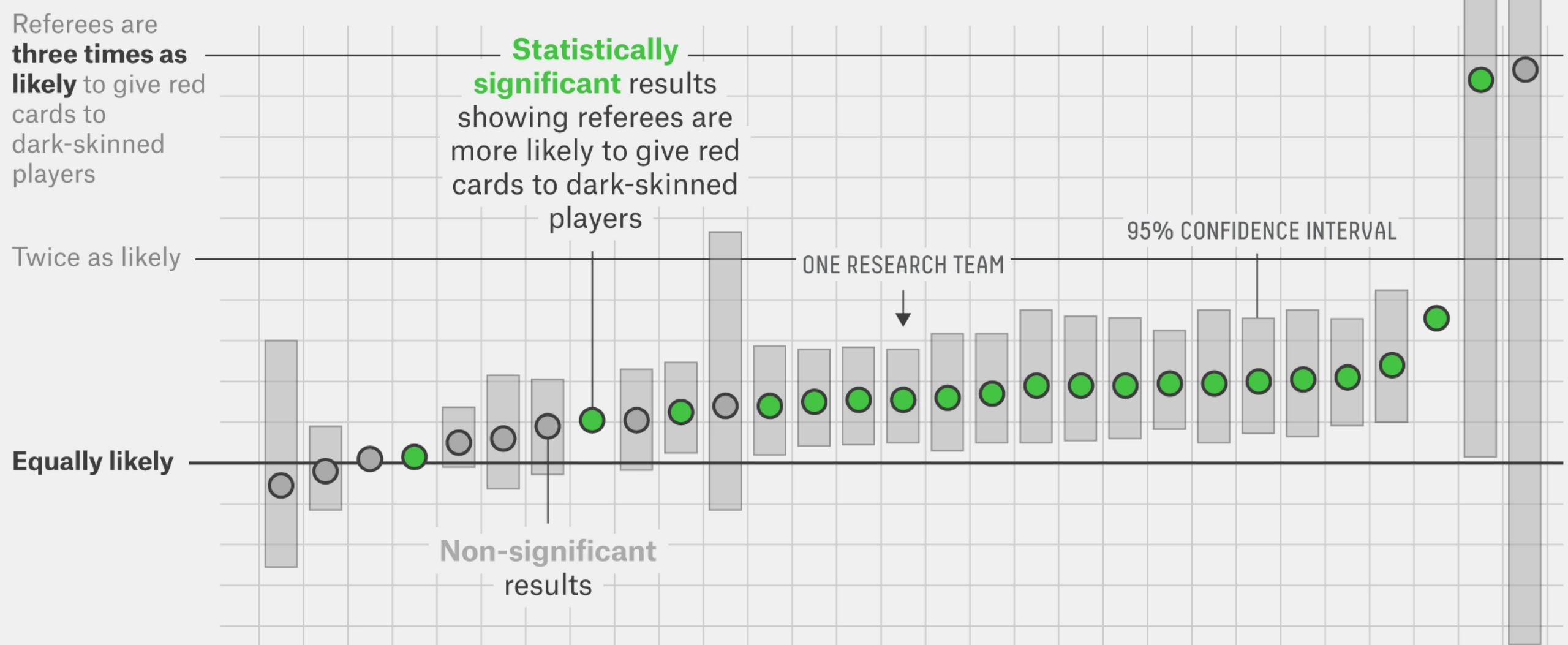
- **variables de contrôle/extérieures** : nous faisons de notre mieux pour que ces variables restent contrôlées et immuables alors que d'autres variables sont modifiées
- **variables indépendantes** : nous contrôlons les valeurs de la variable, et nous soupçonnons qu'elles influent sur les variables dépendantes
- **variables dépendantes** : nous ne contrôlons pas les valeurs, elles sont générées pendant l'expérience et dépendent vraisemblablement de tout le reste

Comment cela se traduit-il dans d'autres ensembles de données?



Same Data, Different Conclusions

Twenty-nine research teams were given the same set of soccer data and asked to determine if referees are more likely to give red cards to dark-skinned players. Each team used a different statistical method, and each found a different relationship between skin color and red cards.



Source: « Science isn't broken – It's just a hell of a lot harder than we give it credit for » ([Christie Aschwanden, 2015](#))

Analytique : immédiate et ciblée



Analytique : la suranalyse immobilise

Nous pouvons être amenés à prendre une décision sans avoir toutes les données en main. Quel est le risque de ne pas prendre de décision par rapport au risque de prendre une décision imparfaite?

L'**immobilisme par la suranalyse** est dû au fait que l'on réfléchit trop à la situation et que l'on s'inquiète du résultat au détriment de la prise de décision. C'est le perfectionnisme, poussé à l'extrême.

« Peu importe la direction dans laquelle vous choisissez de vous déplacer lorsque vous subissez une attaque au mortier, du moment que vous vous déplacez. Les décisions ne sont jamais définitives pour la simple raison que le changement n'est jamais absolu. Au contraire, le changement est permanent. Pour rester concurrentiel et progresser au rythme du changement, il faut des décisions adaptatives qui peuvent être répétées et améliorées au fur et à mesure. » [Jeff Boss, Forbes]

Analytique : éviter l'immobilisme par la suranalyse

1. Soyez-en conscient
2. Établissez le degré de priorité
3. Prenez du recul
4. Demandez des conseils
5. Prenez de petites décisions rapides
6. Fixez un délai
7. Connaissez vos objectifs
8. Limitez l'information que vous recevez

Que visez-vous par l'analyse?

Voulez-vous :

- Agir en fonction de vos données? (peut-être pas de l'analyse?)
- Mieux comprendre quelque chose **en particulier** (des individus bien précis, un groupe en particulier)?
- Arriver à des conclusions **générales** qui vont au-delà du spécifique?

Local c. mondial

Ici c. ailleurs

Passé/présent c. futur

Connaissance de la situation c. préparation aux impondérables

Séquence typique de raisonnement

1. **Commencez par les hypothèses** : connaissances/convictions supposées vraies.
2. **Effectuez le raisonnement.**
3. **Tirez des conclusions** : nouvelles connaissances/convictions potentiellement vraies.

Cette démarche peut aussi servir à générer un **argument logique**.

Vous le faites déjà, inconsciemment

Supposons que je veux descendre un escalier les bras pleins. Le raisonnement se déroule alors **inconsciemment**...

- si je porte trop de choses sur les bras, ALORS je ne peux pas attraper la rampe en descendant l'escalier
- si je ne m'accroche pas à la rampe, alors je risque de trébucher
- si je trébuche, alors je risque de lâcher ce que j'ai dans les bras pour ne pas tomber
- si je laisse tomber mes affaires, alors certaines risquent de se casser
- si mes affaires se cassent, ALORS je serai triste

CONCLUSION : J'ai les bras surchargés.

Vous le faites déjà, inconsciemment

Étant arrivé à cette conclusion, que faire?

- parce que j'ai les bras surchargés, je risque de faire tomber quelque chose et le casser.
- cela me rendrait triste :(
- au lieu de cela, je pourrais faire deux voyages pour pouvoir attraper la rampe
- si je fais deux voyages au lieu d'un, cela ne veut pas dire que je ne laisserai rien tomber et que rien ne se cassera, mais cela augmente ma confiance dans le fait que je ne laisserai pas tomber qqch

Décision et action : Je vais diviser le chargement en deux et faire deux voyages

ou...

non, ça ira, je vais faire attention et faire un seul voyage.

Vulnérabilités du raisonnement

L'**état d'esprit des** complotistes : individus qui sautent à des conclusions non valables parce qu'ils ne peuvent pas raisonner ou déceler les éléments de preuve non concluants (ou parce que leur information de base sont invalides?).

Est-il **plausible** qu'il y ait des puces dans les vaccins contre la COVID? Comment évaluer le degré de plausibilité?

Exercice de réflexion : on vous donne un groupe de logiciens déductifs et un groupe de débatteurs pour vous aider à prendre des décisions. Lequel choisiriez-vous? Est-ce que l'un ou l'autre vous est utile?



Raisonnement formel et rigoureux

Les mathématiciens et les philosophes ont créé des **méthodes formelles** qui introduisent de la **rigueur** dans le raisonnement informel – nous pouvons les considérer comme des outils de raisonnement.

Ces outils augmentent nos chances d'aboutir à **des énoncés vrais**, en lesquels on peut avoir confiance (même si ce n'est pas toujours à 100 %).

Sans rigueur, on risque de se laisser aller à des **raisonnements biaisés**, qui nous empêchent de parvenir à des **conclusions vraies** ou **justifiées**.

Techniques de raisonnement formel

RAISONNEMENT INDUCTIF, PLAUSIBLE,
DÉDUCTIF, ABDUCTIF, ANALOGIQUE



```
graph TD; A[RAISONNEMENT INDUCTIF, PLAUSIBLE, DÉDUCTIF, ABDUCTIF, ANALOGIQUE] --> B[AUTRES TECHNIQUES SPÉCIALISÉES : MÉTHODE SCIENTIFIQUE, RAISONNEMENT STATISTIQUE, MODÉLISATION MATHÉMATIQUE ET INFORMATIQUE]; B --> C[ANALYSE FACTUELLE, QUI PEUT ÊTRE PLUS OU MOINS TECHNIQUE];
```

AUTRES TECHNIQUES SPÉCIALISÉES :
MÉTHODE SCIENTIFIQUE, RAISONNEMENT STATISTIQUE,
MODÉLISATION MATHÉMATIQUE ET INFORMATIQUE

ANALYSE FACTUELLE, QUI PEUT ÊTRE PLUS OU
MOINS TECHNIQUE

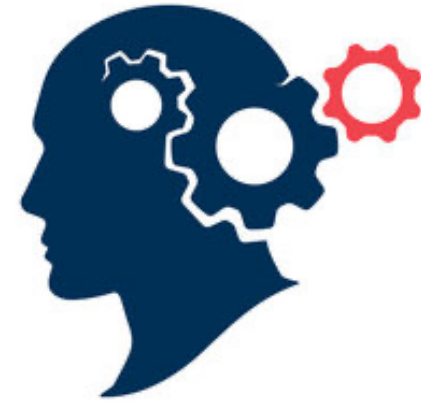
Techniques de raisonnement formel

Stratégies de raisonnement :

- déduire de nouveaux faits en partant de faits existants (raisonnement **déductif**);
- généraliser en partant d'exemples (raisonnement **inductif**);
- raisonner jusqu'à trouver la meilleure explication (raisonnement **abductif**);
- utiliser des analogies et des modèles (raisonnement **analogique**).

Les trois dernières techniques sont des exemples de raisonnement plausible – on n'est pas sûr d'atteindre la vérité, mais on augmente le degré de certitude.

Raisonnement plausible



Considérons le scénario suivant [Jaynes, 2003] :

- vous marchez dans une rue déserte la nuit;
- vous entendez une alarme de sécurité, vous regardez de l'autre côté de la rue, et vous voyez un magasin avec une fenêtre cassée;
- quelqu'un portant un masque sort de la fenêtre cassée avec un sac rempli de téléphones intelligents.

Qu'est-ce que vous en concluez selon le système 1?

Selon le système 2?

Raisonnement plausible

Disons que vous avez conclu que la personne qui sort en rampant du magasin a volé de la marchandise. Comment arrivez-vous à cette conclusion? Ça **ne peut pas** provenir d'une déduction logique fondée sur des indices matériels. En effet,

- la personne qui a rampé hors du magasin **pourrait** être son propriétaire qui,
- en rentrant d'une soirée déguisée, s'est rendu compte qu'il avait égaré ses clés,
- juste au moment où un camion qui passait projetait une brique contre la vitrine du magasin,
- déclenchant l'alarme de sécurité, après quoi
- le propriétaire est entré dans le magasin pour en sortir le contenu avant qu'il ne soit volé,
- et c'est à ce moment que vous êtes arrivé sur la scène.

La conclusion initiale n'est pas **déductive**, mais elle est au moins **plausible**.

Raisonnement déductif et raisonnement plausible

Raisonnement plausible:

Si A est vrai, alors B est plus plausible
 B est vrai

A est plus plausible

Si la « personne est un cambrioleur » (A : vrai), il ne serait pas surprenant de les « voir sortir de la fenêtre cassée avec un sac rempli de téléphones intelligents » (B : plausible).

Vous les observer « sortir de la fenêtre cassée avec un sac de téléphones » (B : vrai).

Vous ne seriez ainsi pas surpris d'apprendre que la « personne est un cambrioleur » (A : plus plausible)

Exercice : Stratégies de raisonnement

Considérez les éléments suivants trouvés dans une note d'information :

- Les sept dernières fois que des pipelines ont été construits dans un territoire occupé par les caribous, leur population y a diminué.
- Des biologistes ont produit une carte montrant les voies de migration du caribou; on en conclut que la pose du pipeline sur le territoire n'entravera pas la migration des caribous.
- Les pipelines n'ont pas affecté les populations d'oies; comme elles et les caribous sont tous deux des animaux sociaux, le pipeline n'affectera pas la population de caribous.
- Les biologistes ont démontré que les caribous n'ont pas peur des objets de grande taille; si les caribous n'ont pas peur, leurs rituels de reproduction ne sont pas affectés; comme les pipelines sont de grands objets, la construction de ce pipeline n'affectera pas les rituels de reproduction du caribou sur le territoire.

Nommez les stratégies appliquées dans chaque raisonnement. Quelle conclusion tireriez-vous de la construction du pipeline et de l'effet qu'il pourrait avoir sur les caribous, par un raisonnement plausible? Que voudriez-vous savoir de plus avant de tirer une conclusion?

Exercice :

Plausibilité

Dans la pièce *Rosencrantz et Guildenstern sont morts* (1966) de Tom Stoppard, les personnages principaux font des paris en lançant une pièce de monnaie. Rosencrantz gagne en obtenant face 92 fois de suite.

Ce résultat n'est bien sûr pas impossible, mais est-il plausible? Si cela vous arrivait, quelle serait votre conclusion?



Machines à raisonner

Un autre avantage de la formalisation du raisonnement est qu'elle nous permet de **l'automatiser**, en le programmant sur ordinateur.

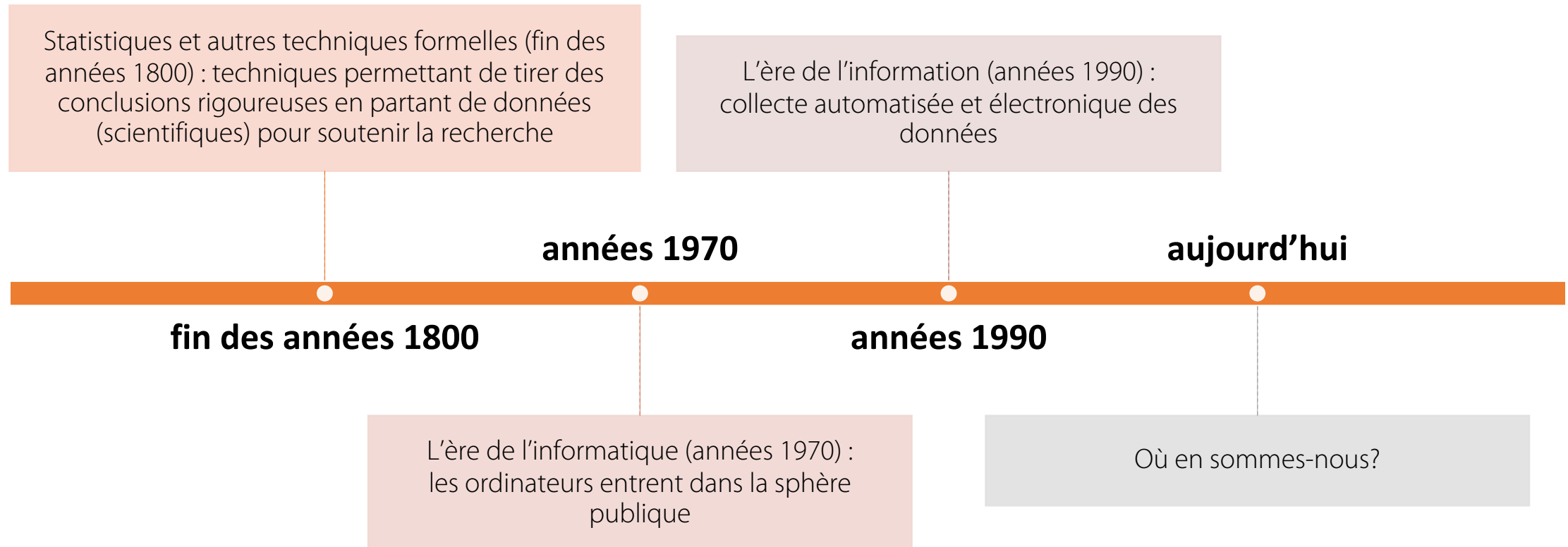
Nous pouvons reformuler le raisonnement comme un processus qui **accepte des entrées** (hypothèses/observations/éléments de preuve) et **produit des sorties** (conclusions).

En automatisant ce processus, nous pouvons faire en sorte que les machines effectuent le raisonnement à notre place.

Le résultat pourrait être plus **fiable**, plus **sûr**, plus **constant**.

MAIS : si on alimente la machine par du n'importe quoi, on obtient en retour du n'importe quoi! À hypothèses boiteuses, conclusions potentiellement fausses!

La montée en force de l'analyse



Analyse à l'ère pré-numérique c. celle à l'ère numérique

Avant :

- L'analyse et les étapes du processus d'analyse pouvaient uniquement être accomplies par des humains.

Aujourd'hui :

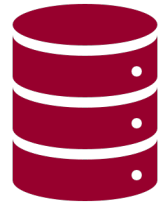
- L'essentiel du processus d'analyse peut être intégré dans un algorithme.
- L'analyse et le processus qui la soutient peuvent être automatisés.
- Nous avons des machines d'analyse.

Avant :

- L'analyse d'une situation donnée se faisait généralement une fois, de façon ponctuelle.
- La personne pouvait effectuer « une analyse » et passer à autre chose.

Aujourd'hui :

- On peut s'attendre à devoir faire des variations de la même analyse à l'infini sur la base des nouvelles données qui nous parviennent régulièrement.



Module 3

Les personnes, l'éthique des données, et la loi



Analyse des données : en quoi cela me concerne-t-il?

En général, même les personnes qui ne pratiquent pas l'analyse doivent être capables de parler aux analystes!

- **ingénieurs de données** : les concepteurs de cuisines, qui doivent savoir ce que...
- **analystes de données** : les cuisiniers vont faire, qui à leur tour doivent savoir ce que...
- **traducteurs de données** : les clients veulent manger, qui pourraient travailler avec...
- **présentateurs de données** : les concepteurs de menus et serveurs, qui veillent à ce que les clients sachent ce qu'ils reçoivent, le tout sous la direction...
- **chef de projet de données** : du chef ou sous-chef, qui s'assure d'approvisionner la cuisine et de bien lier le tout.



Équipes de données et prise de décision

Dans quelle **équipe** êtes-vous?

Les renseignements et les données dont **vous disposez** vous permettent-ils de prendre des décisions **rapides** et **efficaces**?

À quelle fréquence avez-vous besoin du **soutien** de l'équipe chargée des données?

Dans quelle mesure son travail vous permet-il de prendre des décisions rapides et efficaces?

Quels sont les **grands obstacles** qui vous empêchent d'utiliser les données dans votre prise de décision? Quelles sont les autres **difficultés** de la prise de décision fondée sur les données?

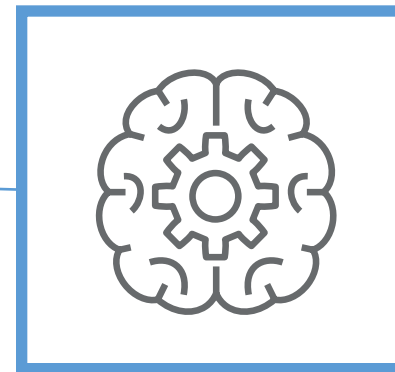
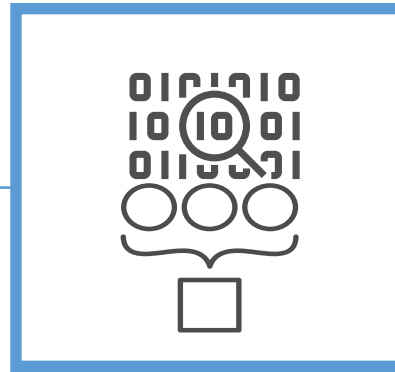
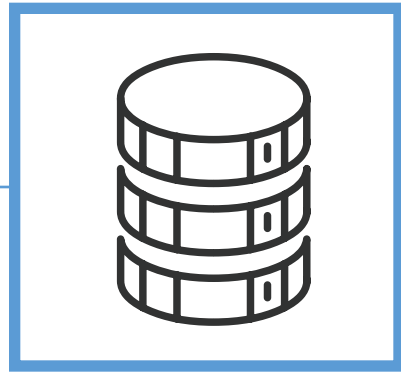
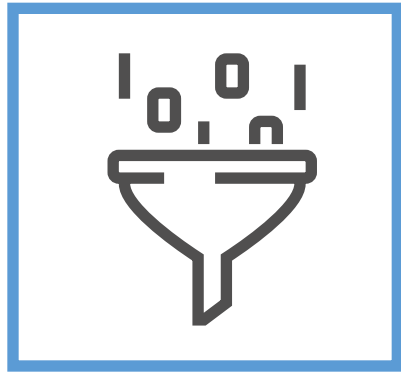
Collecte des données

Stockage des données

Préparation des données

Analyse des données

Présentation des données



L'analyse des données, un sport d'équipe

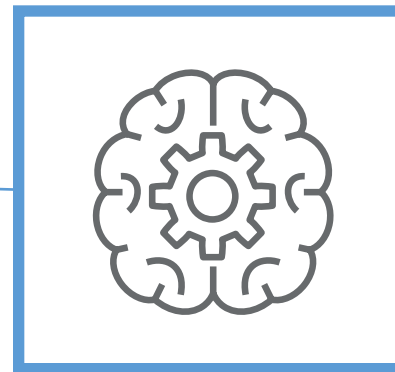
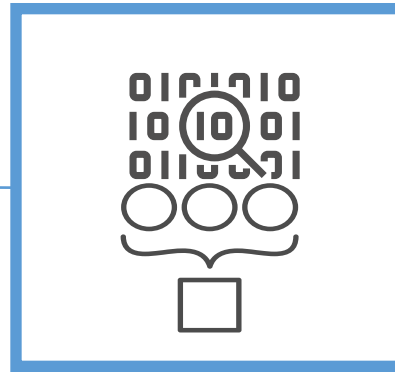
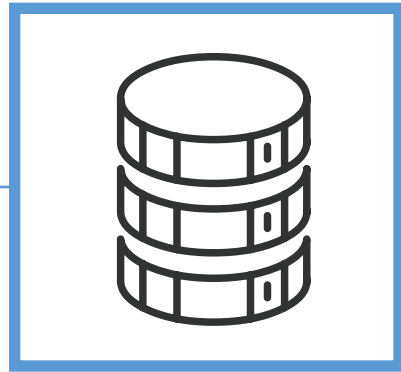
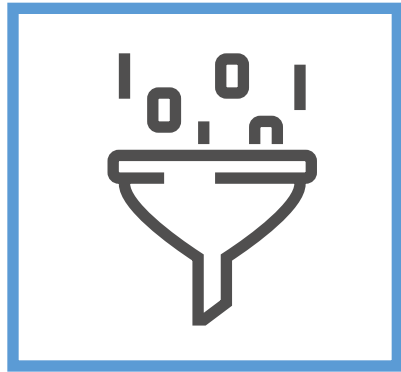
Collecte des données

Stockage des données

Préparation des données

Analyse des données

Présentation des données



Quels rôles soutiennent cette filière?

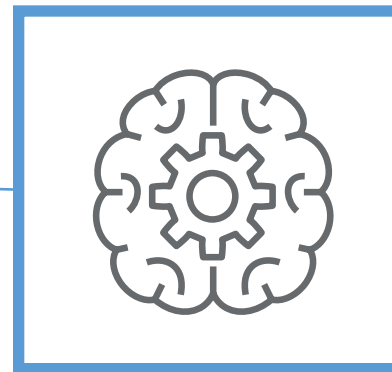
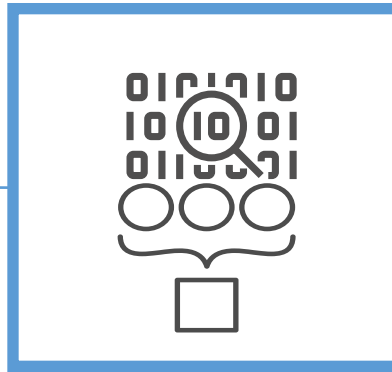
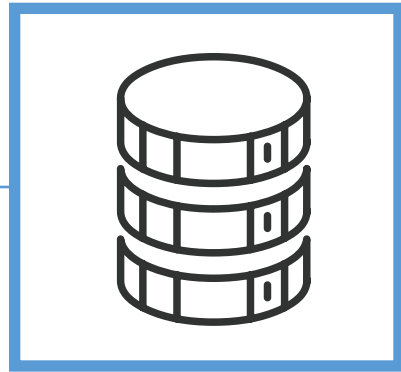
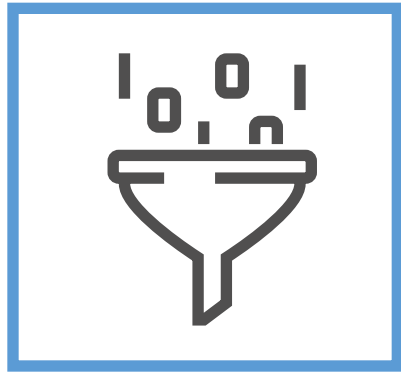
Collecte des données

Stockage des données

Préparation des données

Analyse des données

Présentation des données



Ratio de 1 analyste par équipe de 8 personnes environ

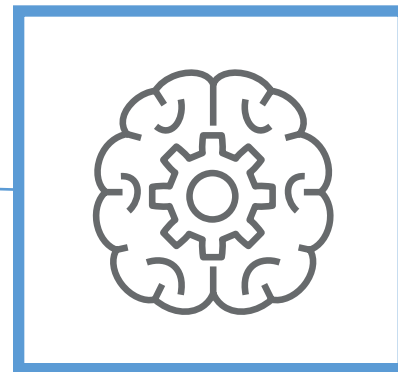
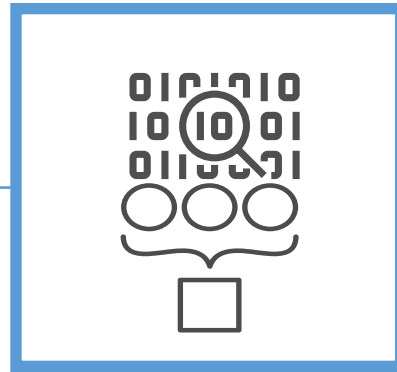
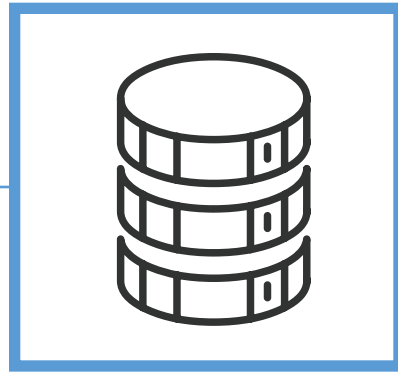
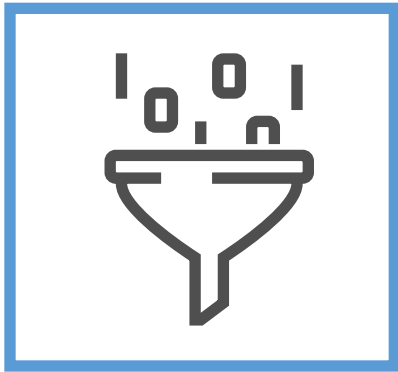
Collecte des données

Stockage des données

Préparation des données

Analyse des données

Présentation des données



Ratio d'environ 1 analyse pour 8 autres activités

Rôles et tâches de l'équipe chargée des données (I)

Ingénierie des données :

conception et mise en œuvre de l'infrastructure de données
- important processus de développement et d'exploitation des TI

Collecte des données :

conception de stratégies de collecte de données et mise en place d'outils de collecte de données

Architecte/gestionnaire de données :

conception et mise en œuvre du stockage et de l'architecture des données

Préparation des données (en collaboration avec la collecte des données) :

les gestionnaires et les analystes de données traitent les données pour les rendre prêtes à être analysées; cette tâche essentielle peut être automatisée, mais doit faire l'objet d'une vérification régulière

Analyse des données : les analystes déterminent la façon de tirer des indications utiles des données et des renseignements; ils conçoivent et mettent en place des algorithmes pour automatiser l'analyse

Rôles et tâches de l'équipe chargée des données (II)

Expert en expérience utilisateur des pipelines de données : conception de l'interface, expérience utilisateur

Communication de données : visualisation des données, présentation des données

Experts en la matière : en savent beaucoup; savent ce qui importe, connaissent les types de données susceptibles d'éclairer la situation, ainsi que la façon d'interpréter et d'appliquer les résultats de l'analyse

Experts en stratégie organisationnelle : ont une vue d'ensemble et savent où l'organisation veut aller; communiquent ce savoir à l'équipe

Chef de projet : garde tout le monde sur la bonne voie et en mode collaboration

Traducteurs de données : connaissent le fonctionnement des différents éléments du pipeline à l'échelle macroscopique, ont une connaissance utile du sujet; mettent les gens en relation et les aident à se parler entre eux

Spécialiste :

connaissance approfondie d'un sujet

Généraliste :

vaste connaissance d'un sujet

Amateur :

passionné

Professionnel :

payé et censé assurer un certain rendement, et un certain degré de compétence (obligations)



Considérons l'analyse de données sur ordinateur comme une activité « **semi-professionnelle** »



On peut toujours former une équipe d'amateurs ou de semi-professionnels



Il peut s'agir d'un arrangement moins formel, mais les différents rôles peuvent toujours entrer en jeu

Quelques analogies utiles

	Médecine	Cuisine	Automobile
Amateur	Tout le monde Secouriste	Cuisinier à domicile	Propriétaire du véhicule Passionné de voitures
Semi-pro	Ambulancier paramédical	Vendeurs de pâtisserie à but non lucratif	Pilote semi-professionnel
Professionnel	Directeur d'hôpital Infirmier Médecin : généraliste, spécialiste	Propriétaire de restaurant Chef cuisinier Chef pâtissier	Mécanicien de station-service Mécanicien de garage Carrossier

Quelle est la place des membres de votre équipe?

Questions préliminaires :

1. Quelle est la partie du pipeline qui vous attire le plus?
2. Préférez-vous concevoir ou exécuter ce qu'un autre a conçu? Ou les deux?
3. Êtes-vous une généraliste (vue d'ensemble) qui aime savoir un peu de tout, ou une spécialiste (soucieuse du détail)?
4. Savez-vous écrire des programmes informatiques/des scripts (ou voulez-vous être capable de le faire)?
5. Avez-vous une formation en mathématiques ou en statistiques (STIM)?
6. Aimez-vous travailler avec les technologies de l'information?
7. Aimez-vous faciliter la communication entre les membres d'une même équipe?
8. Avez-vous une connaissance approfondie des activités de votre organisation ou des sujets traités?
9. Avez-vous une connaissance approfondie des objectifs de l'organisation? Aimez-vous la stratégie?

Généraliste : polyvalent, communique sans frontières





	Collecte de données	Stockage de données	Préparation de données	Analyse	Présentation	Sujet	Organisation
Amateur	exécution	exécution			expérience utilisateur	au début	au début
Semi-pro	conception du questionnaire	architecture de données (conception)		statistiques	conception exécution	après le début du projet	après le début du projet
Professionnel		architecture pour l'apprentissage automatique	non structuré, semi-structuré	apprentissage automatique			

apprentissage non supervisé, ARS

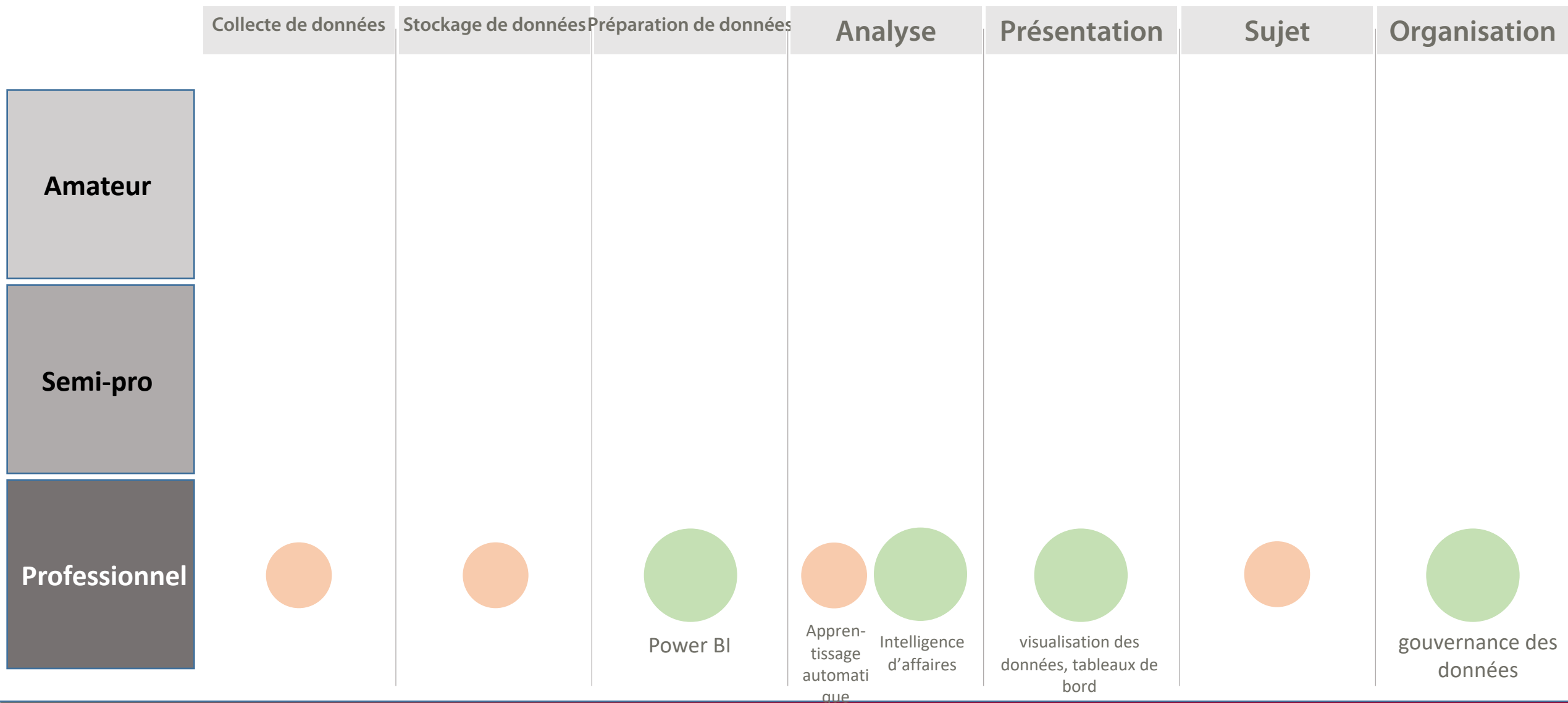
Hybride : éventail restreint de tâches liées aux données (autres compétences dans des domaines autres que les données)

	Collecte de données	Stockage de données	Préparation de données	Analyse	Présentation	Sujet	Organisation
Amateur							
Semi-pro					raconter des histoires	L'IA en tant que service, facteurs humains	
Professionnel				analyse des besoins		questions juridiques et GC	gestion












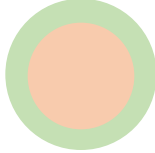

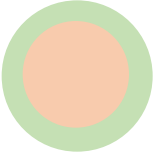
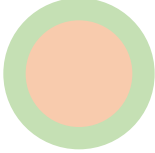
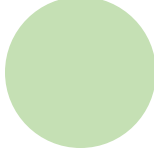
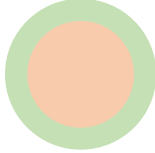
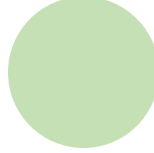
Spécialiste : ciblé, peut parler à certains publics

	Collecte de données	Stockage de données	Préparation de données	Analyse	Présentation	Sujet	Organisation
Amateur							
Semi-pro							
Professionnel	 échantillonnage d'enquêtes, moissonnage du Web		 traitement des données, nettoyage des données	 Analyse exploratoire des données, statistiques, algorithme de reconnaissance d'objets, apprentissage automatique	 visualisation des données		

Hybride : large éventail de tâches liées aux données



Les lacunes ne sont pas idéales, bien sûr, mais l'absence de lacunes ne garantit pas la réussite.

	Collecte de données	Stockage de données	Préparation de données	Analyse	Présentation	Sujet	Organisation
Amateur							
Semi-pro							
Professionnel							

Exercice : Quelle est votre place (ou celle de votre équipe)?

	Collecte de données	Stockage de données	Préparation de données	Analyse	Présentation	Sujet	Organisation
Amateur							
Semi-pro							
Professionnel							

Expérience ou nouvelle perspective

Partir de zéro (**équipe rouge**) OU faire appel à des solutions toutes faites/personnes du domaine (**équipe verte**)

Nouvelles perspectives

Points de vue biaisés sur les données

Pas d'ancrage analytique

Susceptible au « déjà vu, déjà fait »

Détection naturelle des anomalies

Certains domaines peuvent être occultés par des hypothèses préétablies

Vous pouvez vérifier trois fois vos hypothèses et vos méthodes en faisant appel à l'**équipe rouge** pour remettre en question les conclusions perçues.

Récolter les fruits peut prendre longtemps.

Différenciation : personnes chargées de l'informatique et des données

TI

Serveurs/infonuagique
Connexion aux données
Accès et autorisation

Données

Environnements de code
Dépôts de code
Hiérarchies des projets
Progiciels à code source libre
Modèles de production et analyse

Critères de recrutement de « personnes » spécialisées dans les données

Curieuse

Disposée à apprendre

Communicative

Résolue

Bien informée

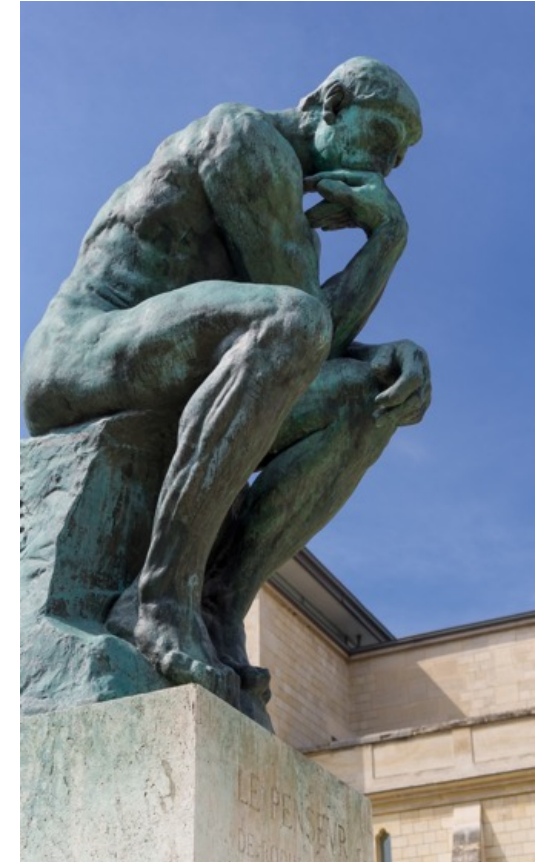
Esprit d'analyse

Qu'est-ce que l'éthique?

De manière générale, l'éthique fait référence à l'**étude** et à la **définition** des **bons** et **mauvais** comportements.

Nous avons tous une éthique personnelle, n'est-ce pas?

- être honnête
- être juste
- être objectif
- être responsable
- être compatissant?
- etc.





Prochain exercice : Le voile de l'ignorance

La situation de chacun nous rend tous subjectifs, alors comment peut-on se mettre d'accord sur les règles qui régissent le fonctionnement du monde?

On doit s'imaginer assis derrière un voile d'ignorance qui nous empêche de savoir qui nous sommes et d'établir des liens avec notre propre situation.

Étude de cas

Votre entreprise est toujours à la recherche des meilleurs talents, notamment pour les postes techniques.

La politique de l'entreprise soutient la diversité et l'inclusion.

L'embauche prend beaucoup de temps et vous craignez que les préjugés des membres du jury n'influencent les décisions.

Avec l'aide d'une équipe d'IA exceptionnelle, vous automatisez ce processus.

Les processus assistés par l'IA trouvent un % élevé de personnes talentueuses, qui s'intègrent à la culture organisationnelle et qui aiment leur travail.

Étude de cas : l'IA d'embauche d'Amazon

Vous êtes plus susceptible d'être embauché si vous vous appelez Jared et que vous jouez à la crosse.

L'IA se comportait de manière partielle, en ne recommandant pas l'embauche de femmes.

N'étant pas sûre de pouvoir supprimer le biais ou de détecter les comportements partiels à l'avenir, Amazon a décidé d'interrompre le projet.

Parlons de ce cas (et des autres études de cas).

Amazon ditched AI recruiting tool that favored men for technical jobs

Specialists had been building computer programs since 2014 to review résumés in an effort to automate the search process



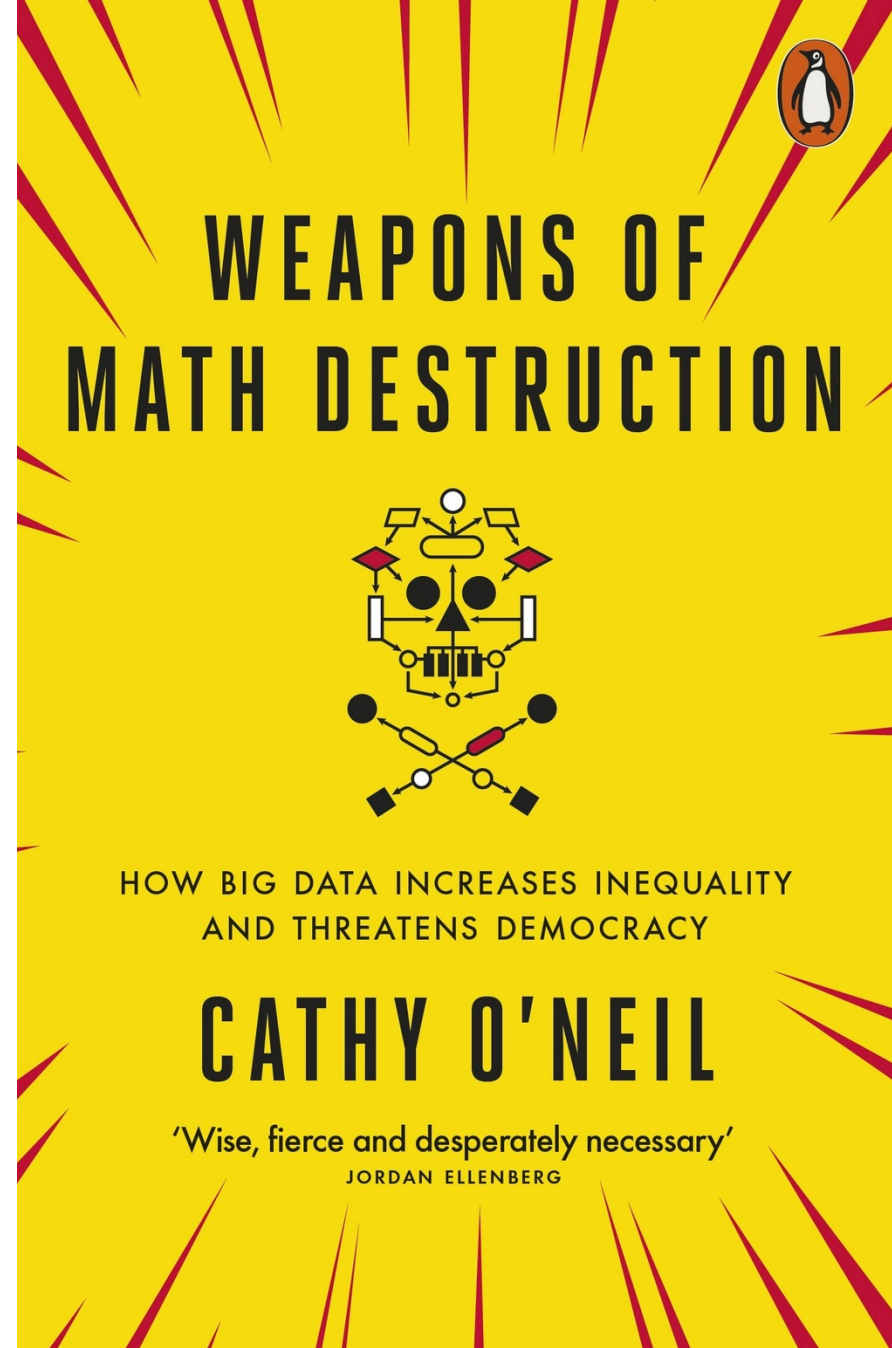
▲ Amazon's automated hiring tool was found to be inadequate after penalizing the résumés of female candidates. Photograph: Brian Snyder/Reuters

La menace

Dans son livre sur le pouvoir des données, M^{me} Cathy O'Neil présente plusieurs exemples qui servent de mise en garde.

« Un programme informatique peut parcourir rapidement des milliers de CV et les classer dans des listes bien ordonnées. Non seulement cette caractéristique permet de gagner du temps, mais elle aussi été présentée comme une façon de procéder juste et objective. Après tout, il ne s'agissait pas d'humains pleins de préjugés traitant des montagnes de papier, mais de machines traitant froidement des chiffres. » [TRADUCTION]

Les applications mathématiques à l'origine de l'économie des données reposent sur des choix effectués par des êtres humains faillibles. Il n'y a aucun doute que certains choix ont été faits avec les meilleures intentions. Néanmoins, nombre de ces modèles et algorithmes ont encodé, dans les logiciels qui gèrent de plus en plus notre vie, des préjugés, des malentendus et des partis pris bien humains. »



L'éthique dans le contexte des données

Questions relatives à l'éthique des données :

- Qui, le cas échéant, est propriétaire des données?
- Y a-t-il des limites à l'usage qu'on peut faire des données?
- Certaines analyses comportent-elles des biais de valeur?
- Y a-t-il des catégories à éviter dans l'analyse des données personnelles?
- Faut-il mettre certaines données à la disposition de tous les chercheurs?

Y a-t-il des leçons à tirer des principes de PCAP® des Premières Nations :

- **propriété, contrôle, accès, possession.**



Exercice : Le voile de l'ignorance

Supposons ce qui suit :

- une banque a l'obligation d'accroître sa valeur pour ses actionnaires;
- elle envisage d'appliquer un nouveau processus décisionnel fondé sur l'IA dans les demandes de prêt, et d'éviter ainsi au personnel d'avoir à accomplir des tâches fastidieuses et réduire les défauts de paiement;
- la banque a accès à des données exclusives et publiques sur les demandeurs de prêts.

Sans savoir la position que vous occupez dans la société (par exemple, si vous travaillez pour la banque ou si vous demandez un prêt), avez-vous des questions sur cette proposition? Quels principes éthiques entreraient en jeu?

Pratiques exemplaires

« **Ne pas nuire** » : les données recueillies auprès de l'individu ne doivent pas être utilisées pour lui nuire.

Consentement éclairé :

- L'individu doit autoriser la collecte et l'utilisation de ses données.
- L'individu doit bien comprendre ce qu'il autorise, et les conséquences éventuelles pour lui et pour autrui.

Respecter la « vie privée » : excessivement difficile à assurer à l'ère de la recherche constante de données personnelles sur Internet.

Pratiques exemplaires

Garder les données publiques : les données doivent rester publiques (toutes? la plupart? aucune?).

Possibilité d'accepter ou de se retirer : le consentement éclairé exige la possibilité de se retirer.

Anonymiser les données : suppression des champs d'identification des données avant l'analyse.

D'autres idées?

Nouvelles tendances qui apparaissent dans le droit

Canada

- GC : faire l'[évaluation de l'incidence algorithmique](#) avant la production de n'importe quel système de décision automatisé.
- Commissaire à la protection de la vie privée (Loi sur la protection des renseignements personnels et les documents électroniques)
Modification proposée :
 - Définir les systèmes de décision automatisés comme étant toute technologie qui assiste ou remplace le jugement des humains.
 - Il faut expliquer aux gens la prédiction ou recommandation, et la manière dont leurs renseignements personnels ont été utilisés.

Europe

- Règlement général sur la protection des données
 - Article 22 : droit de ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé (sauf exception).
 - Article 15 : en cas de prise de décision automatisée, droit d'obtenir des informations utiles concernant la logique sous-jacente.

Questions juridiques

Profilage :

- utilisez-vous des données personnelles pour tirer des conclusions injustes, non éthiques ou discriminatoires?

Surveillance :

- les gens sont-ils placés dans une parade d'identification sans fin?

Responsabilité :

- êtes-vous responsable de ce que fait l'IA?

Étude de cas : Programmes d'IRCC

Immigration, Réfugiés et Citoyenneté Canada (IRCC) cherche très activement des moyens d'améliorer ses services par un usage judicieux de l'automatisation et de l'IA.


Dans un projet pilote récent, des techniques d'automatisation et d'apprentissage automatique ont été utilisées pour accélérer le traitement de certaines demandes.

Planning highlights

Client experience, automation and experimentation

In recent years, the Department has been working to advance its network of online tools to support both clients and staff and improve services, particularly as they relate to application processing times. These efforts will continue in 2018–2019 across key business lines as the





En fin de compte, il ne s'agit pas (seulement) d'analyse!

Dans un cadre professionnel, l'analyse **ne se fait pas dans le seul but de faire une analyse.**

Sur le terrain, l'analyse soutient les objectifs de l'organisation.

- **Analogie de la cuisine** : dans le monde de la restauration, le client est roi.
- **Analogie de la passion pour les voitures** : certaines personnes font de la mécanique par plaisir, d'autres travaillent sur les données par plaisir, mais en fin de compte, il s'agit de savoir qui est le propriétaire/conducteur de la voiture et de connaître l'usage qu'il compte en faire.

Ne perdez pas de vue vos objectifs ultimes.

Qui sait quels sont les objectifs ultimes? Divulgâcheur : **ce n'est pas nécessairement vous!**



Module 4

L'intelligence économique et l'analytique

Qu'entendons-nous par "intelligence" ?

Analyse des données

Généralement considéré comme une sous-composante de l'analytique des données.

Différentes méthodes utilisées pour fournir une vue historique des données

Processus d'observation des données, puis d'organisation/de visualisation pour en extraire des informations utiles.

Analytique des données

Portée plus large que l'analyse des données

Comprend des méthodes permettant d'inclure des prédictions futures

Comprend diverses méthodes utiles à la gestion des données

Intelligence économique

Utilise les résultats de l'analytique et des analyses

Utilise la vue du passé et les prédictions futures pour créer une connaissance situationnelle afin de prendre des décisions.

Utilise les méthodes d'analyse et d'analytique dans un contexte commercial.



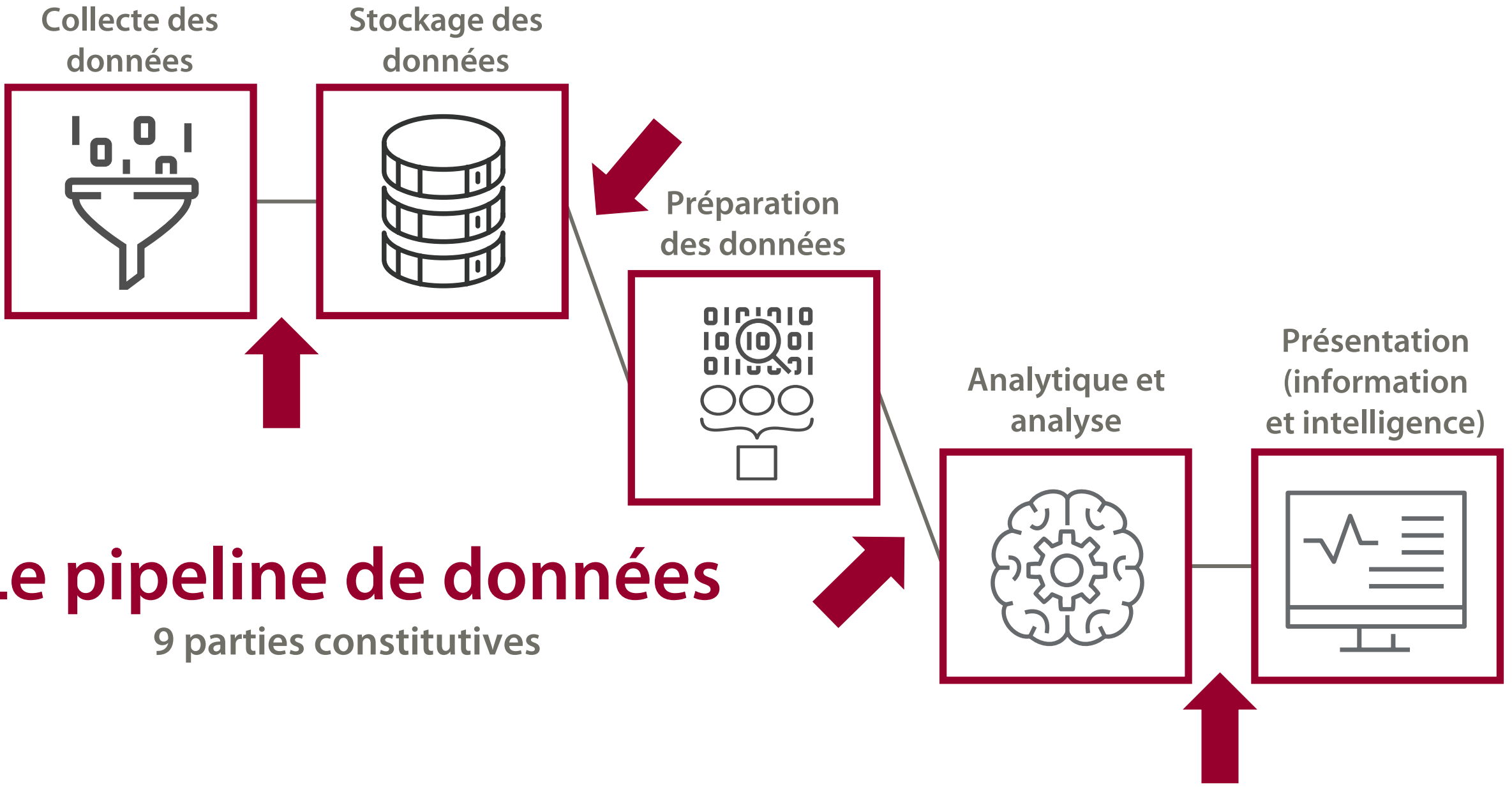
Qu'entendons-nous par "analytique" ?

Nous utilisons souvent l'analyse et l'analytique de manière interchangeable, mais l'analyse est généralement considérée comme une **sous-composante** de l'analytique.

Nous pensons qu'il s'agit de méthodes qui nous fournissent des informations utiles.

Mais l'analytique en est venue à représenter une **large gamme d'outils de traitement des données**, couvrant l'entreposage de données, la gestion de l'information d'entreprise, la gestion de la performance d'entreprise et la gouvernance.

Certaines "méthodes" fournissent des informations/intelligences ; d'autres "méthodes" transforment, nettoient, manipulent et stockent les données (**pipeline de données**).



Le pipeline de données

9 parties constitutives

Analytique des données

On parle souvent de l'analytique en termes de "**modes**".

Certains modes peuvent être plus "**valables**" que d'autres en fonction du contexte.

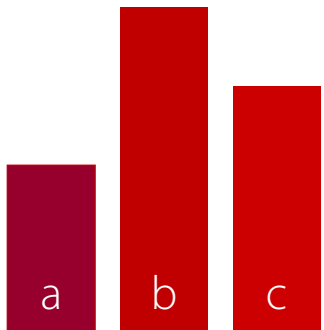
Nous pourrions ne **jamais être en mesure d'atteindre certains modes** (par exemple, prédictif) en raison de problèmes avec nos données.

Ces modes **dépendent de la technologie**, les modes à valeur élevée nécessitant des outils et des méthodes plus sophistiqués.

Modes analytiques

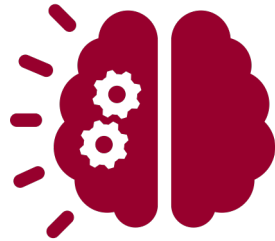
L'analyse peut être décomposée en quatre **modes principaux** :

Descriptive



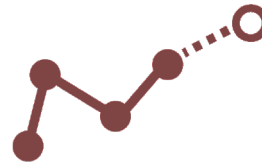
Montrer **ce qui s'est** passé

Diagnostic



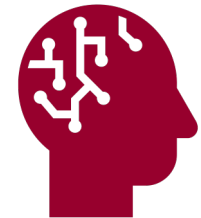
Expliquer **pourquoi** quelque chose s'est produit

Prédictif



Devinez **ce qui va** se passer

Prescriptive



Suggérer **ce qui devrait** se passer

Faible valeur
Faible difficulté



Valeur élevée
Difficulté élevée

Modes analytiques

Descriptive : Rapport budgétaire fournissant un instantané à un moment donné.

Diagnostique : Pourquoi mon budget est-il inférieur à ce que je pensais initialement ?

Prédictif : Combien vais-je avoir dans mon budget au prochain trimestre ?

Prescriptive : Comment dois-je modifier mon profil de dépenses pour éviter le déficit ?

Analyse descriptive

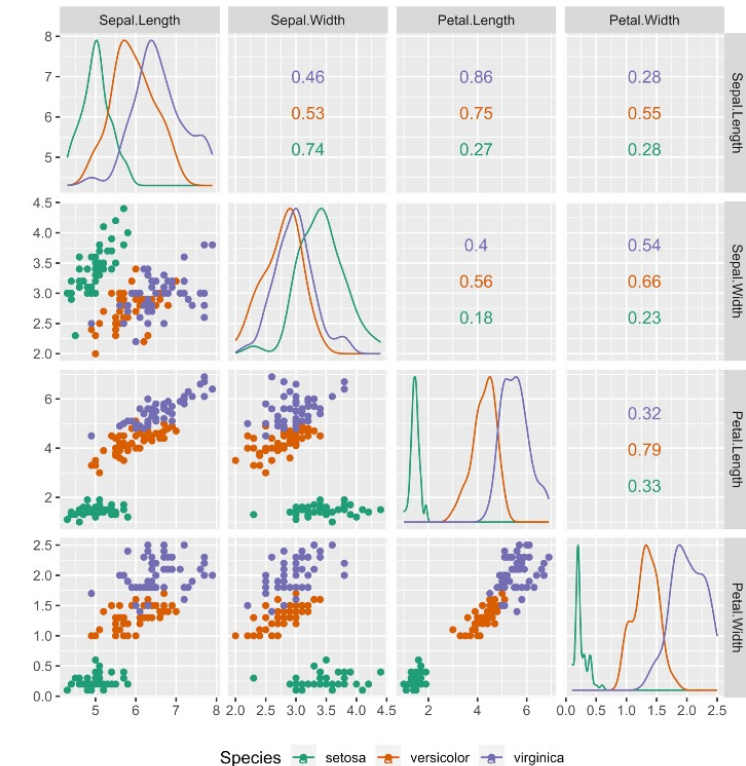
Examine les données de manière statistique pour montrer ce qui s'est passé **dans le passé**.

Aide une entreprise à comprendre sa performance **en fournissant du contexte** pour aider les parties prenantes à interpréter les informations.

Généralement, sous forme **de graphiques, de tableaux, de rapports, de tableaux de bord**, etc.

L'intelligence est généralement retardée (les données ne sont généralement **pas actuelles**) et se concentre **sur un sous-ensemble ou un seul ensemble de données**.

Très souvent, les ensembles de données ne sont **pas appropriés/suffisamment grands** pour déduire une signification statistique ou pour effectuer des tests fiables.



Analyse diagnostique

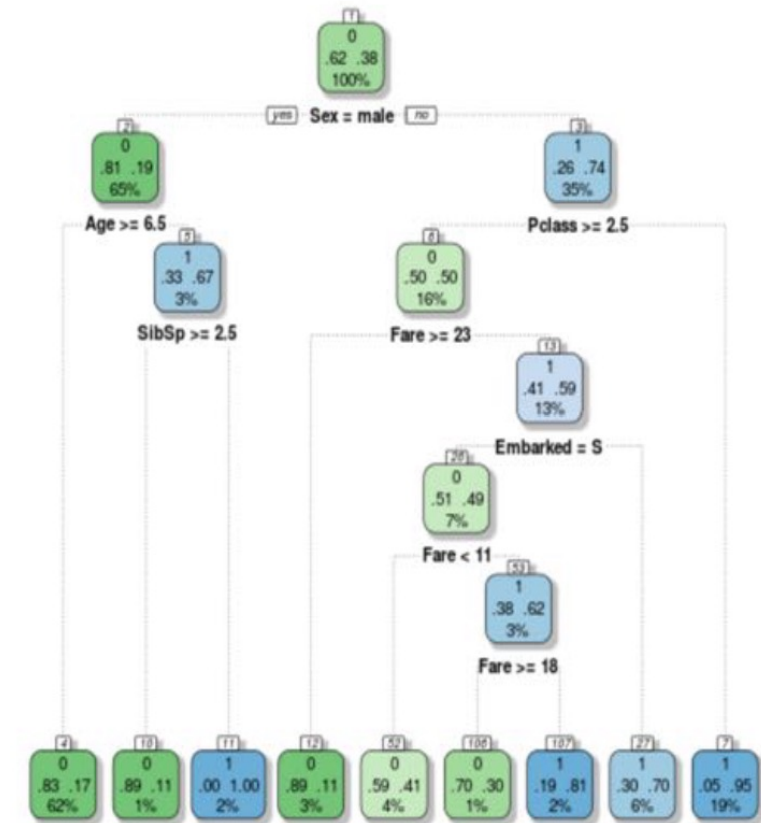
Fournit une analyse plus approfondie pour répondre à la question : **pourquoi cela s'est-il produit ?**

Utilise très souvent des outils **d'analyse des causes profondes formalisés** (Taproot, Causal Tree, etc.)

Difficile à automatiser et sujet à la variabilité humaine

L'intelligence **confond souvent corrélation et causalité**

Si les ensembles de données sont appropriés, l'IA/ l'apprentissage automatique peut apporter une **valeur significative** (par exemple, les arbres de décision).



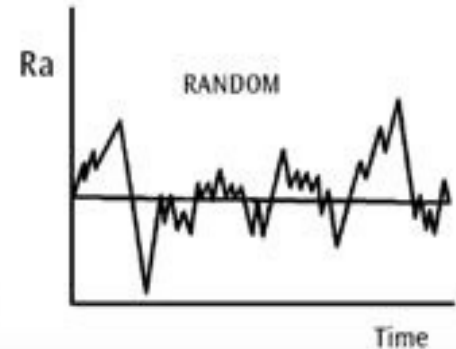
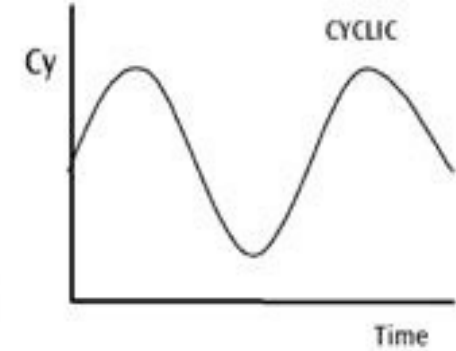
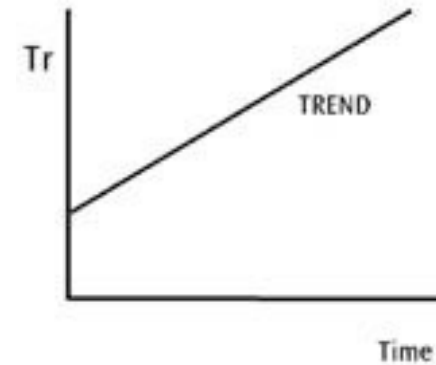
Analyse prédictive

Applique diverses méthodes (statistiques, ML, etc.) pour **modéliser et extraire des caractéristiques** à partir de données antérieures.

Applique des modèles et des schémas aux données **actuelles** pour **prédire** ce qui pourrait se passer.

L'intelligence peut souffrir de la **dérive du modèle**, d'un mauvais choix des données d'entraînement pour la modélisation et de nombreux autres problèmes (notamment la taille de l'ensemble de données).

Très populaire, mais facile à rater



Analyse prescriptive

Tire parti d'approches qui peuvent suggérer **diverses pistes d'action**

L'intelligence est très souvent **probabiliste** et donc potentiellement **subjective** (65% de chance de pluie, donc on prend un parapluie ?)

L'intelligence est souvent utilisée dans le cadre de la **prise de décision augmentée**

Le Saint-Graal de l'analytique

		Prisoner 2	
		Cooperate	Defect
Prisoner 1	Cooperate	3, 3	0, 5
	Defect	5, 0	1, 1

Qu'est-ce que l'intelligence économique ?

De nombreuses définitions, dont la plupart sont au mieux incohérentes et au pire contradictoires.

Nous devrions comparer avec d'autres formes de collecte de renseignements, par exemple :

- champ de bataille/combat
- diagnostic (médical)
- réponse aux catastrophes
- sports

Qu'est-ce que l'intelligence économique ?

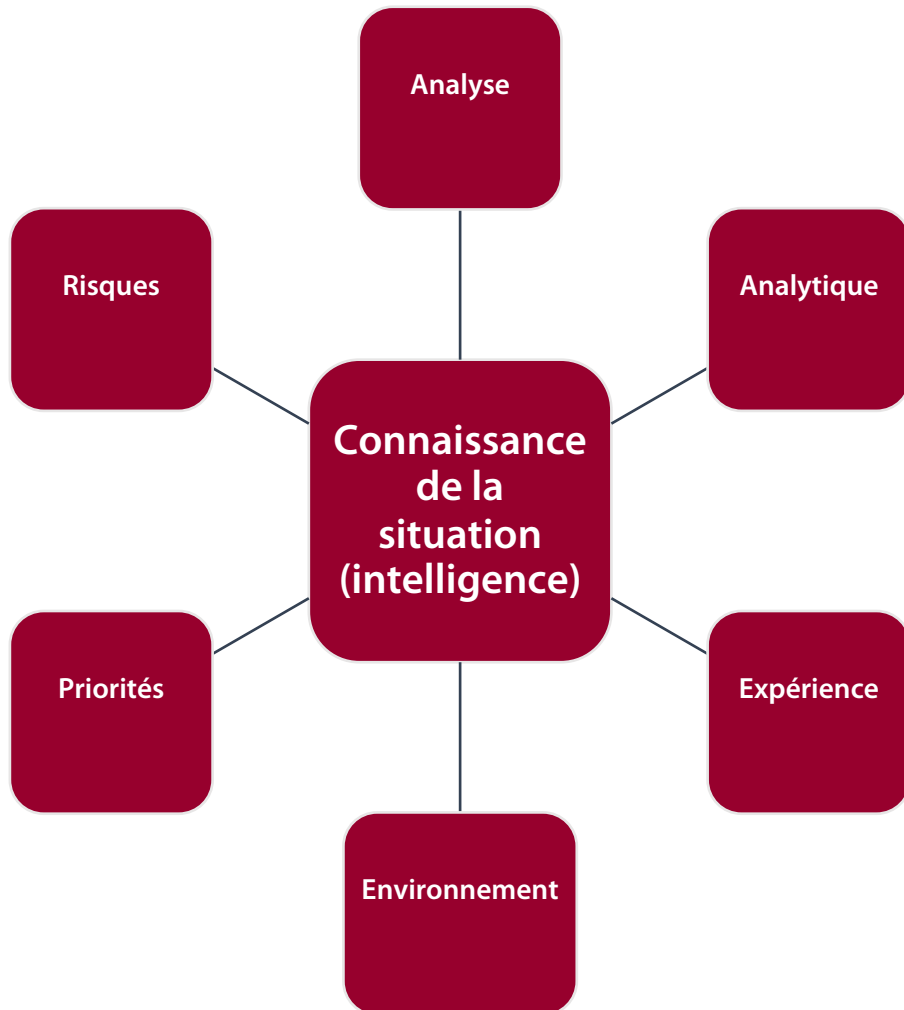
Dans tous les cas, l'actrice (commandante, médecin, coordinatrice, joueuse, etc.) a besoin d'une **connaissance précise de la situation** pour prendre des décisions éclairées et opportunes.

Dans tous les cas, la qualité de la connaissance de la situation s'améliore avec la technologie.

Par conséquent, le terme « Business Intelligence » a évolué pour représenter une gamme de technologies qui aident les **décideurs** au sein d'une "entreprise".

La connaissance de la situation reste l'objectif principal, quelle que soit la technologie.

Qu'est-ce que l'intelligence économique ?



Catégorie	Exemple de sport
Analyse	L'équipe locale a gagné 60% des derniers matchs contre le même adversaire
Analytique	Les bookmakers prévoient une victoire à une cote de 3:1 pour l'équipe locale.
Expérience	Tous les membres de l'équipe > 2 ans de ligues majeures
Environnement	Jouer à domicile
Priorités	Match de pré-saison, le résultat ne compte pas dans le championnat.
Risques	85% de chance de pluie

Histoire de l'intelligence économique

Fin des années 1800 : les gens ont commencé à reconnaître qu'ils pouvaient utiliser les données pour obtenir un avantage concurrentiel.

1865

1950s : l'avènement de la première base de données d'entreprise pour l'aide à la décision

1950s

Années 1980-1990 : les ordinateurs et les données deviennent de plus en plus disponibles - entrepôts de données, exploration de données - mais ils restent très techniques et spécialisés.

1980-90s

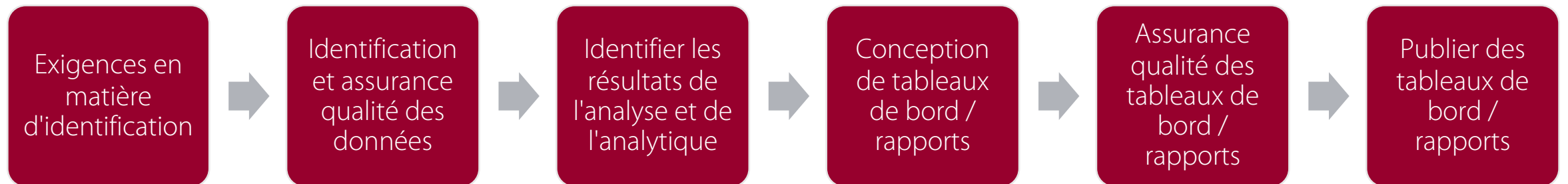
Années 2000 : tentative de faire passer l'analyse d'entreprise des mains des spécialistes de l'extraction de données à celles des experts du domaine.

2000s

2020s : big data, techniques spécialisées, tableaux de bord, logiciel en tant que service

2020s

Processus d'intelligence économique



Améliorer les prévisions

Nous avons vu dans nos exemples de superprévision que les gens peuvent s'améliorer dans l'aspect prédiction (prévision) de l'intelligence économique.

Quelles sont les stratégies ou techniques centrées sur l'humain pour améliorer cette compétence ?

Deux possibilités :

- **Pré-mortem**
- **Rétrospective ('Backcasting')**



Exercice : Pré-mortem

Une analyse post-mortem permet de déterminer les causes d'un mauvais résultat, à une petite exception près : le patient est déjà mort.

Pré-mortem : imaginez-vous à un moment donné dans le futur, après avoir échoué à atteindre un objectif, et regardez en arrière pour savoir comment vous êtes arrivé à cette destination - c'est une analyse post-mortem avant la mort du patient.

Étapes:

1. identifier l'objectif à atteindre ou la décision envisagée
2. choisissez un échéancier pour atteindre cet objectif.
3. Imaginez que nous sommes le lendemain de la date limite et que vous regardez en arrière à partir de cette date ; donnez 5 raisons "sous votre contrôle" et 5 raisons "hors de votre contrôle" pour lesquelles les choses n'ont pas fonctionné.

Exercice: Rétrospective (*Backcasting*)

Rétrospective (*Backcasting*) : imaginez-vous à un moment donné dans le futur, ayant réussi à atteindre un objectif, et regardez en arrière pour savoir comment vous êtes arrivé à cette destination.

Le processus est semblable à celui des pré-mortems (mais les directions changent).

Avantages (pré-mortem/rétrospective) :

- obtenir une vue de l'extérieur
- La sagesse de la foule
- éviter la pensée de groupe (*groupthink*)
- élimine le risque que les gens "ne veulent pas être la roue grinçante".

Table d'exploration des décisions

Pré-mortem

Rétrospective

3 raisons "sous votre contrôle"

1.		1.	
2.		2.	
3.		3.	

3 raisons "hors de votre contrôle"

1.		1.	
2.		2.	
3.		3.	

S'adapter à la malchance et l'atténuer

Après un prémortem/retrospective (*backcasting*) :

1. modifiez votre décision en fonction des nouvelles connaissances acquises ; augmentez les chances que les bonnes choses se produisent et réduisez les chances que les mauvaises choses se produisent ;
2. chercher des moyens d'atténuer l'impact de la malchance ; couverture (qui réduit l'impact de la malchance lorsqu'elle se présente) ; il y a un prix, et nous espérons ne jamais avoir à y recourir (par exemple, l'assurance contre les inondations).

L'Alliance rebelle a réussi à voler les plans de l'Étoile de la mort. L'Empire n'a pas entrepris un prémortem : "toute attaque menée par les Rebelles contre cette station serait un geste inutile, quelles que soient les données techniques dont ils disposent". Les Rebelles ont appris qu'une torpille frappant un petit orifice d'échappement externe pouvait provoquer une réaction en chaîne fatale.

C'est un cas classique d'excès de confiance.

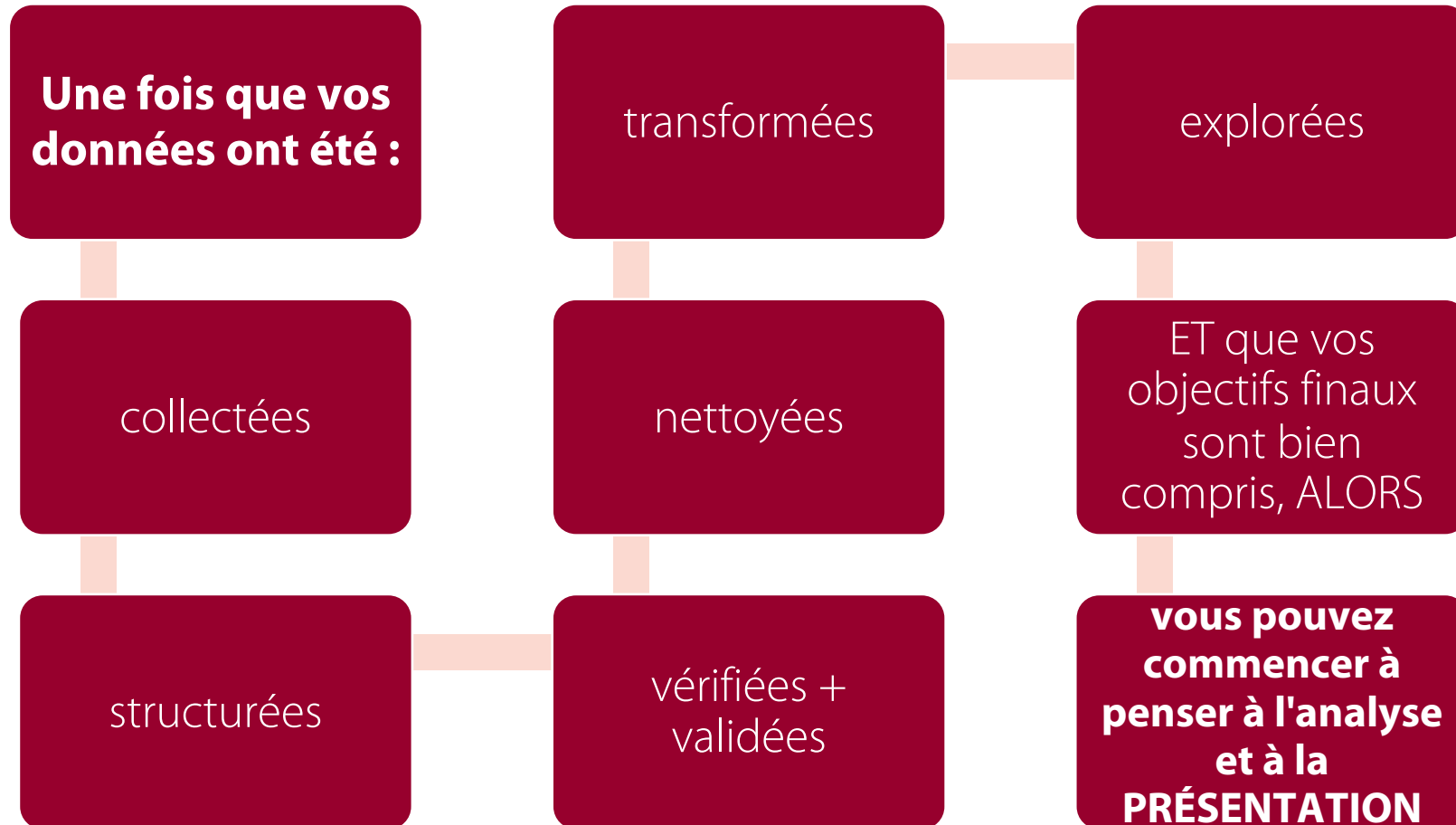
Que se passe-t-il après l'analyse ?

Un résultats d'analyse n'est aussi utile que la façon dont il est communiqué.

Nous ne saurions trop insister sur ce point.



Sommes-nous déjà arrivés ?



Persuasion

Vous avez pris une décision fondée sur des données ou des preuves : comment convaincre les gens que c'était la bonne ?

Nous avons parlé de description, d'explication, de prédiction, de prescription.

Sans persuasion, cela pourrait s'avérer être une perte de temps.

Dans un contexte commercial, il ne suffit pas de savoir (ou de soupçonner) que quelque chose est ou sera le cas, il faut **communiquer** et **convaincre** d'autres personnes.



Récit de base (*story spine*) : PIXAR

Once upon a time there was _
Every day, _.
One day _.
Because of that, _.
Because of that, _.
Until finally _.

The Story Spine

Exercice : construisez une base narrative (*story spine*) relative à une décision réussie prise par votre organisation.

Est-ce tout ce qu'il y a à dire sur le récit de données ?

Risques liés au récit

"Ouvrez n'importe quel journal, regardez n'importe quelle émission d'information télévisée, et vous trouverez des experts qui prévoient ce qui va arriver. Certains sont prudents. La plupart sont audacieux et confiants. Une poignée d'entre eux prétendent être des visionnaires olympiens capables de voir des décennies dans l'avenir. À quelques exceptions près, ils ne sont pas devant la caméra parce qu'ils possèdent un quelconque talent de prévisionniste.

La précision est rarement mentionnée. Vieilles prévisions, vieilles nouvelles - vite oubliées. Le seul talent indéniable des présentateurs est leur capacité à raconter une histoire convaincante avec conviction, et cela suffit. Nombre d'entre eux se sont enrichis en colportant des prévisions dont la valeur n'a pas été testée auprès de dirigeants d'entreprises, de fonctionnaires et de personnes ordinaires qui ne songeraient jamais à avaler des médicaments dont l'efficacité et la sécurité sont inconnues, mais qui paient régulièrement pour des prévisions aussi douteuses que des élixirs vendus à l'arrière d'un chariot." [Tetlock]



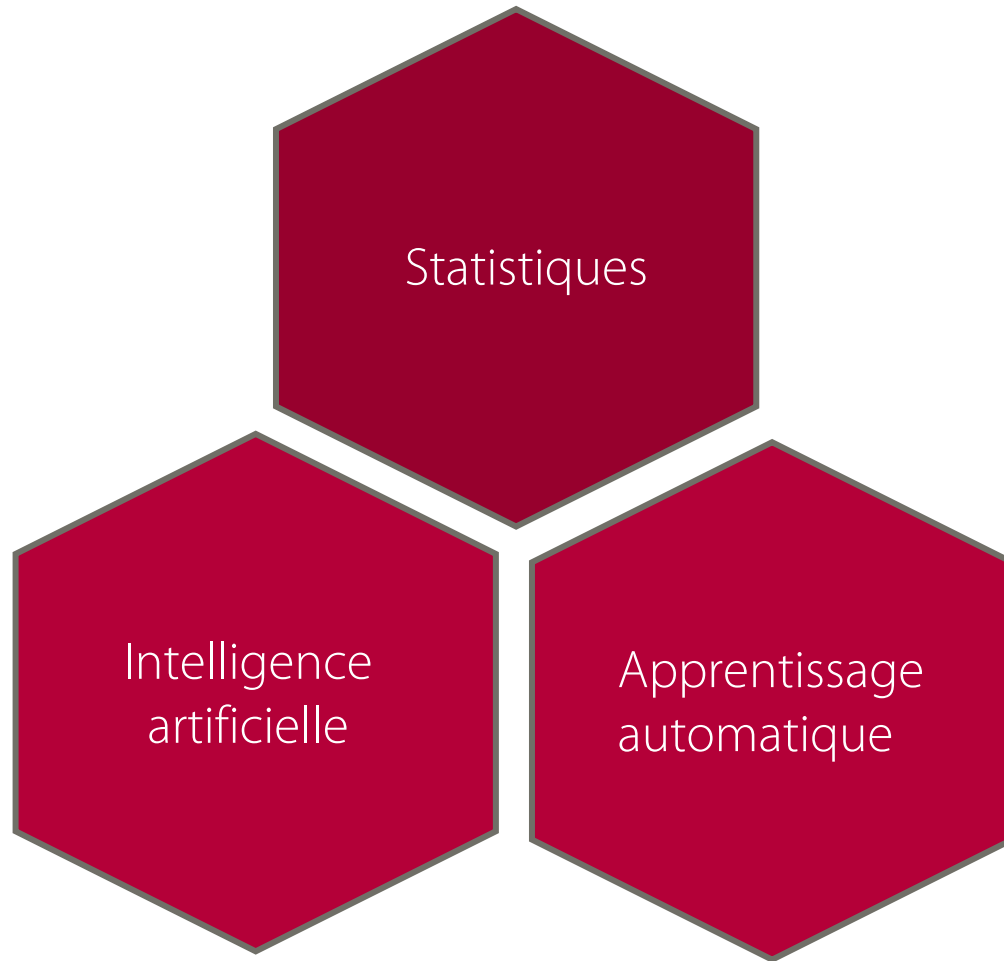
Module 5

L'analyse et l'appui décisionnel

Modèle de maturité des données

Niveau de maturité des données	Description	Critères probables de prise de décision
0	Aucune donnée	L'instinct
1	Pas assez de données	Extractions manuelles
2	Trop de données	Adopter une plateforme BI
3	Analyse	Plate-formes multiples
4	Apprentissage	Commencer à comprendre la prédiction
5	Action	Prédictif/prescriptif

Concepts et méthodes de haut niveau



Extraction de règles d'association, regroupement

Régression et classification

Analyse de séries chronologiques

Détection des anomalies et analyse des valeurs aberrantes

Exploration de textes et analyse de sentiments

Traitement du langage naturel

Apprentissage par renforcement

Systèmes de recommandation

Analyse des réseaux (sociaux)

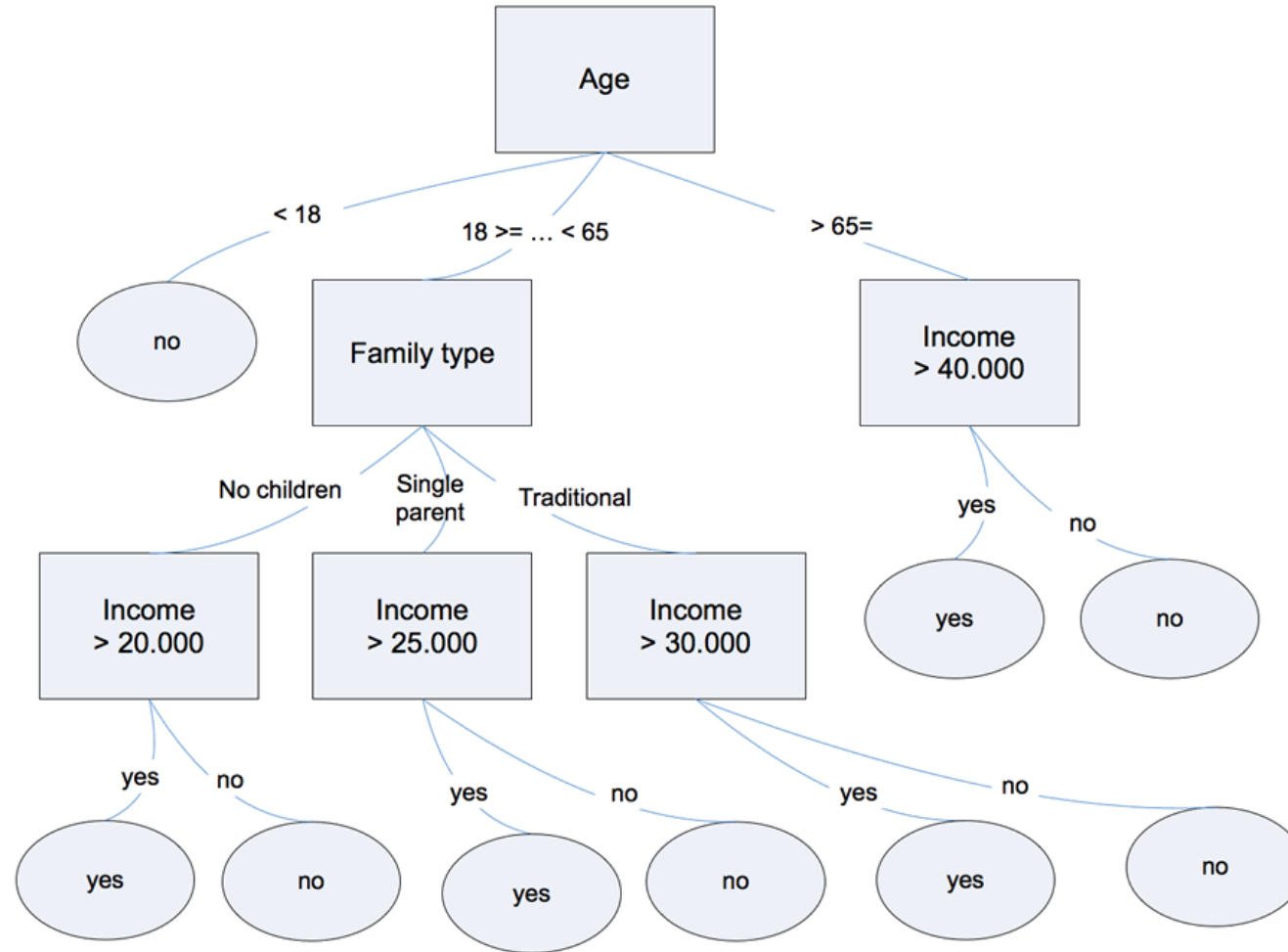
Méthodes Monte Carlo et OR

Flux de données, etc.

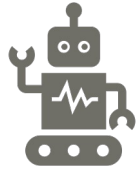
Ramper



Marcher



Courir



La robotique : fusionner logiciel et matériel



Apprentissage automatique et IA

Supervisé	Non supervisé	Renforcement
Classification	Regroupement	Q-Learning
Régression	Recommandeurs	SARSA
Apprentissage profond	Règles d'association	Réseau Deep Q

Défis et obstacles

Données erronées/non représentatives

Surdimensionnement/sous-dimensionnement

Big data et goulots d'étranglement technologiques

Perfectionnisme/manque de rigueur

Poser les mauvaises questions/ne pas poser de questions

Interprétabilité et capacité d'action

Toutes les décisions ne doivent pas être prises sur la base de l'analytique.

Conseils pratiques

Écoutez vos analystes !

Méfiez-vous de la tyrannie des succès passés

Ne vous attendez pas à des miracles

L'analytique guide le processus de décision, elle ne le remplace pas

Apprenez à échouer de manière utile/intelligente

L'analyse n'est pas (seulement) une question d'outils, d'infrastructure et de méthodes (mais c'est un élément important).

Moins, c'est plus : gardez les choses simples (lorsque c'est approprié).

Applications possibles

Déterminer la probabilité qu'une personne ait besoin d'un logement subventionné.

Déterminer quelles missions diplomatiques sont "similaires" les unes aux autres

Prévoir le temps d'attente moyen dans les aéroports canadiens pendant la journée

Déterminer la probabilité d'échec d'un projet en fonction de divers facteurs

Prédire si un conteneur entrant dans le pays contient des produits illégaux

Prévoir la probabilité de défaillance d'un dépôt de déchets nucléaires sur de longues périodes.

Suggestions (communication à l'interne)

1

Créez une feuille de route pour les objectifs d'analyse

2

Veillez à ce que les principaux indicateurs clés de performance soient communiqués

3

Évaluez les aspects organisationnels par rapport aux aspects techniques

4

Énoncez vos principales hypothèses

5

Établissez des relations solides avec les parties prenantes

6

Construisez des visualisations en fonction de l'audience

7

Formulez des personas ou des récits d'utilisateurs

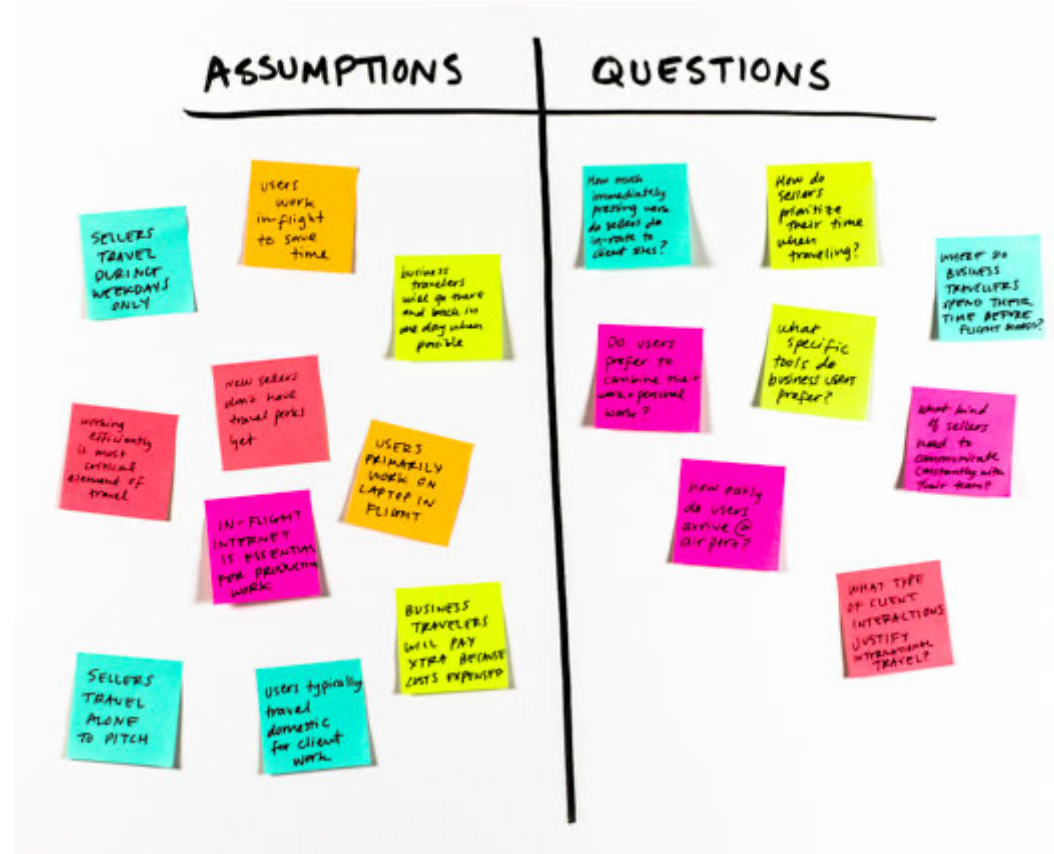
8

Rédigez des histoires captivantes

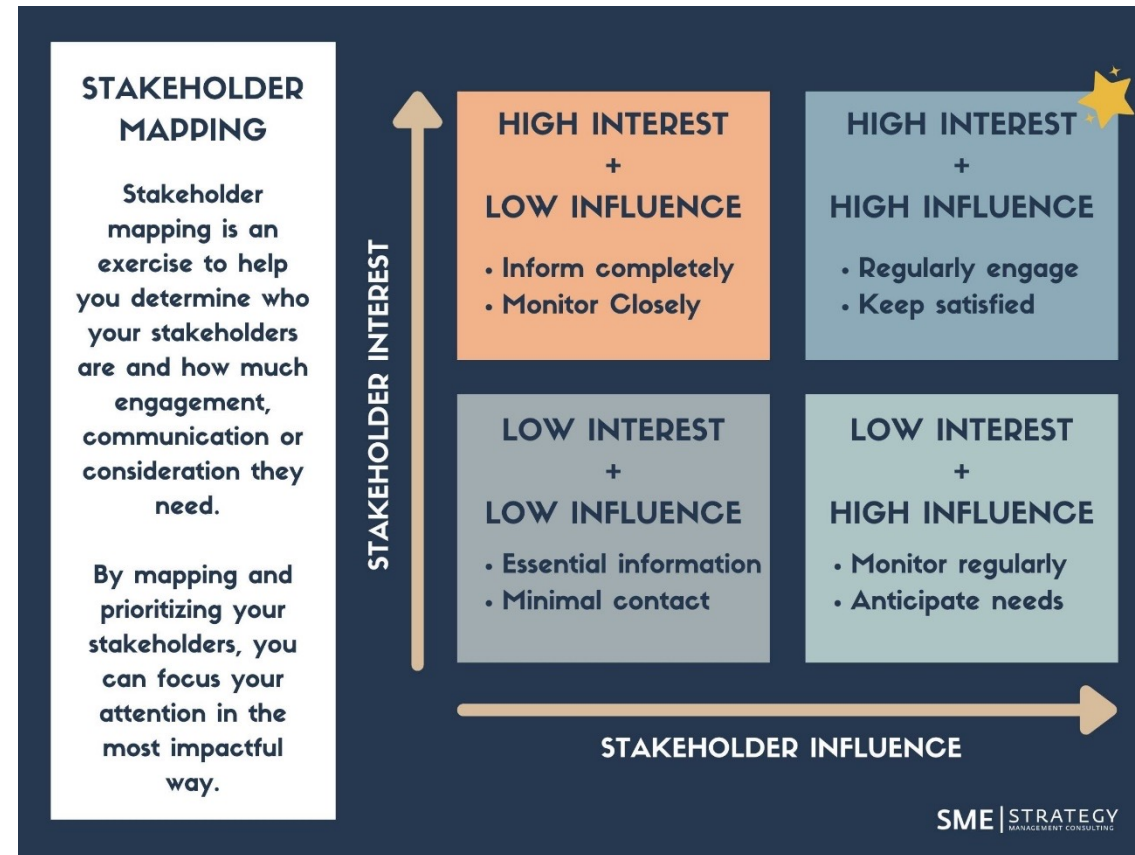
Évaluez les aspects organisationnels et techniques



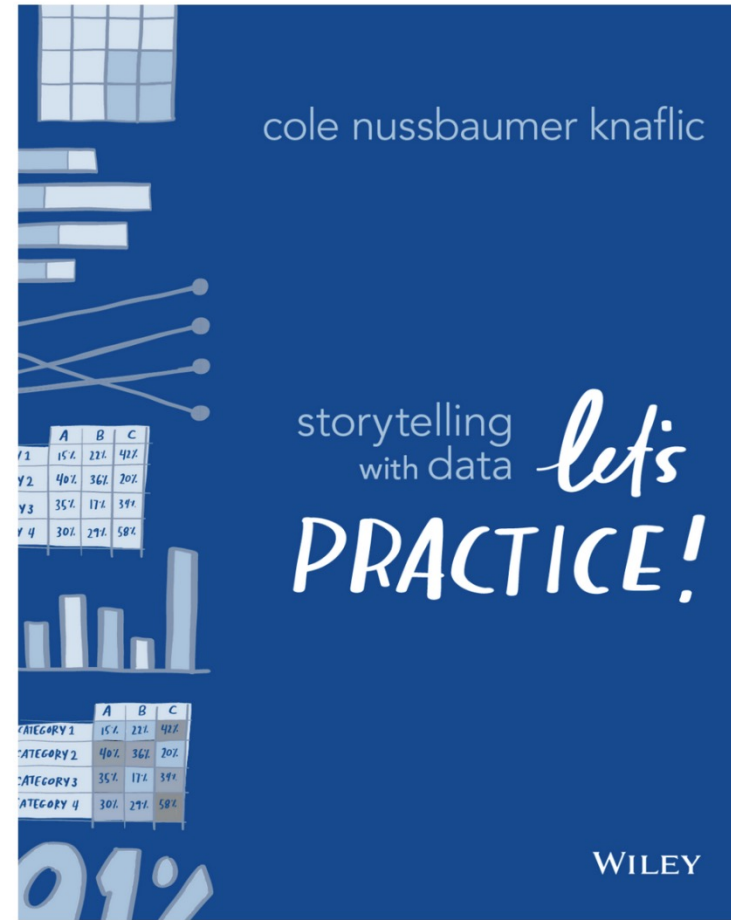
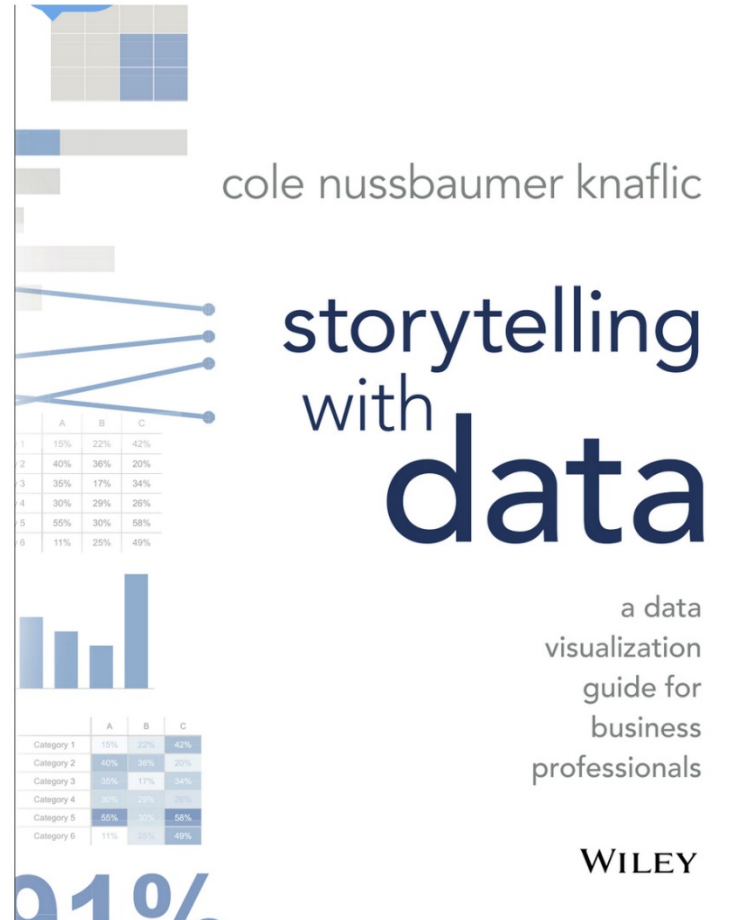
Énoncez vos principales hypothèses



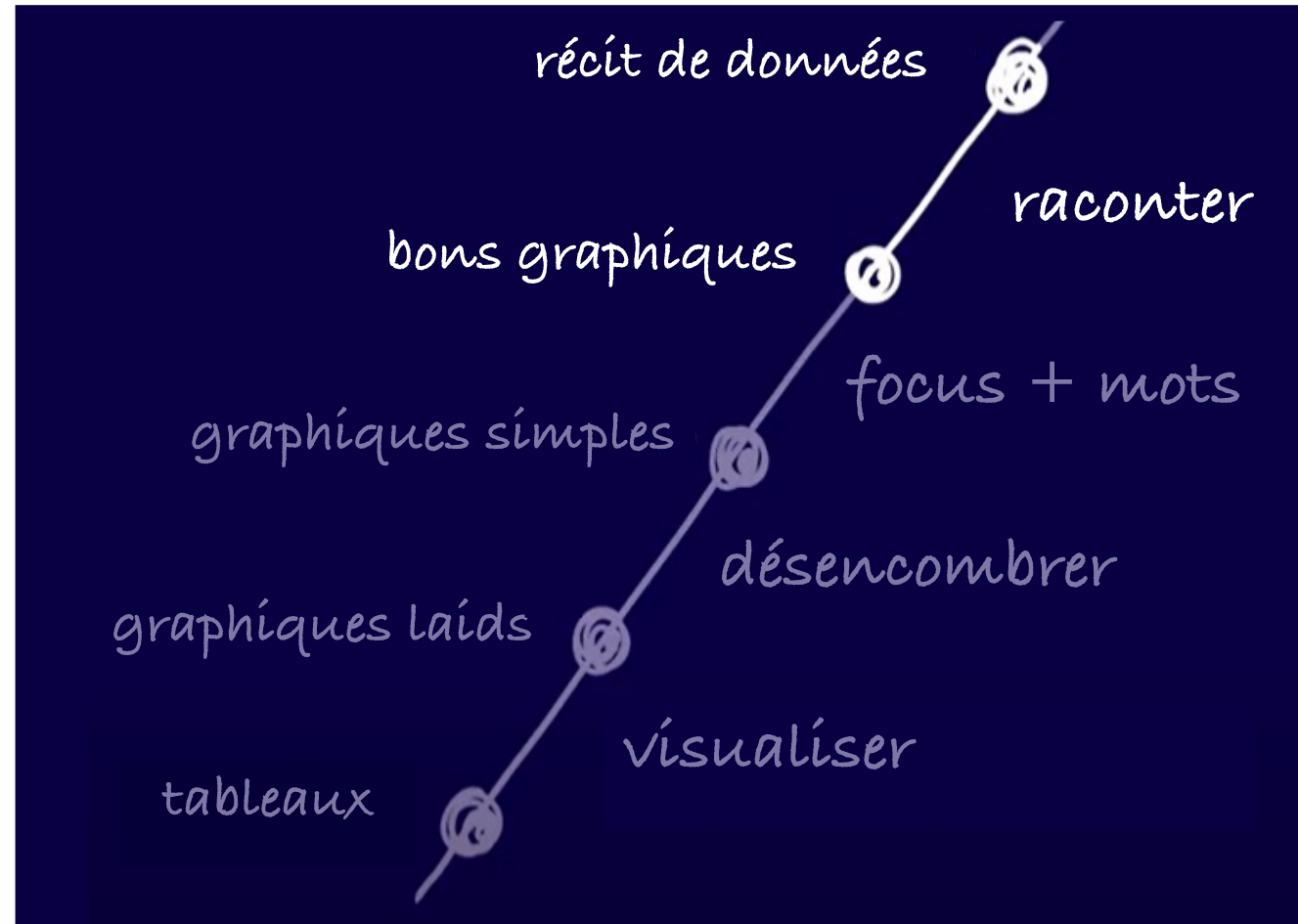
Établissez des relations avec les parties prenantes



Créez des histoires captivantes



Évolution d'une visualisation



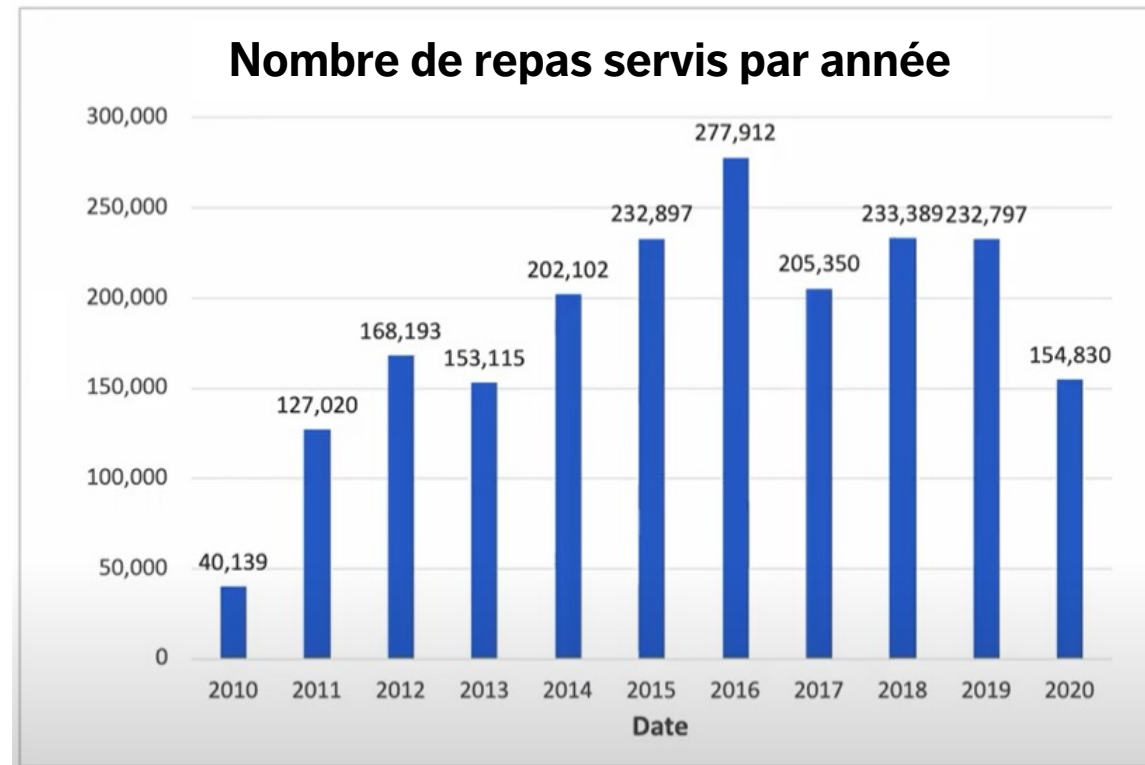
Source: Storytelling with

Évolution d'une visualisation

Meals served over time

Campaign Year	Meals Served
2010	40,139
2011	127,020
2012	168,193
2013	153,115
2014	202,102
2015	232,897
2016	277,912
2017	205,350
2018	233,389
2019	232,797
2020	154,830

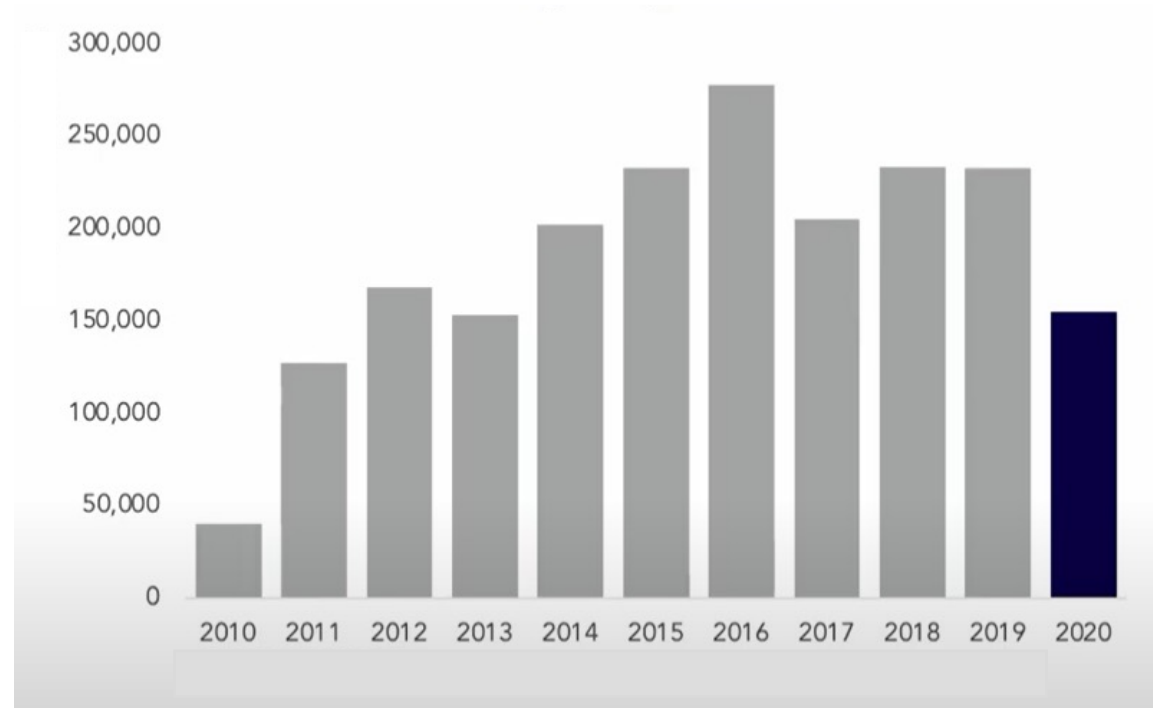
Évolution d'une visualisation



Source: Storytelling with Data

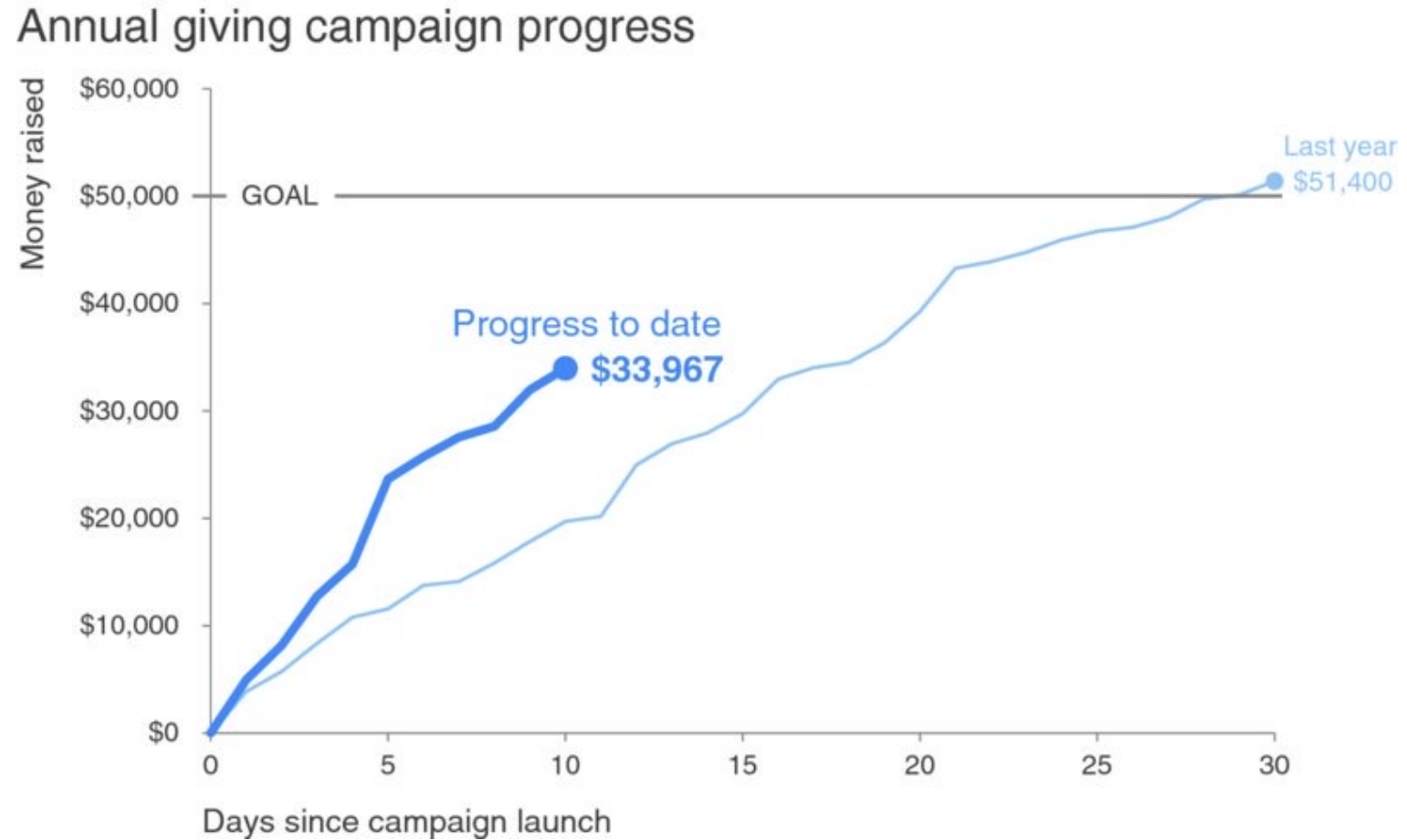
Évolution d'une visualisation

Nombre de repas servis par année: **baisse importante en 2020**



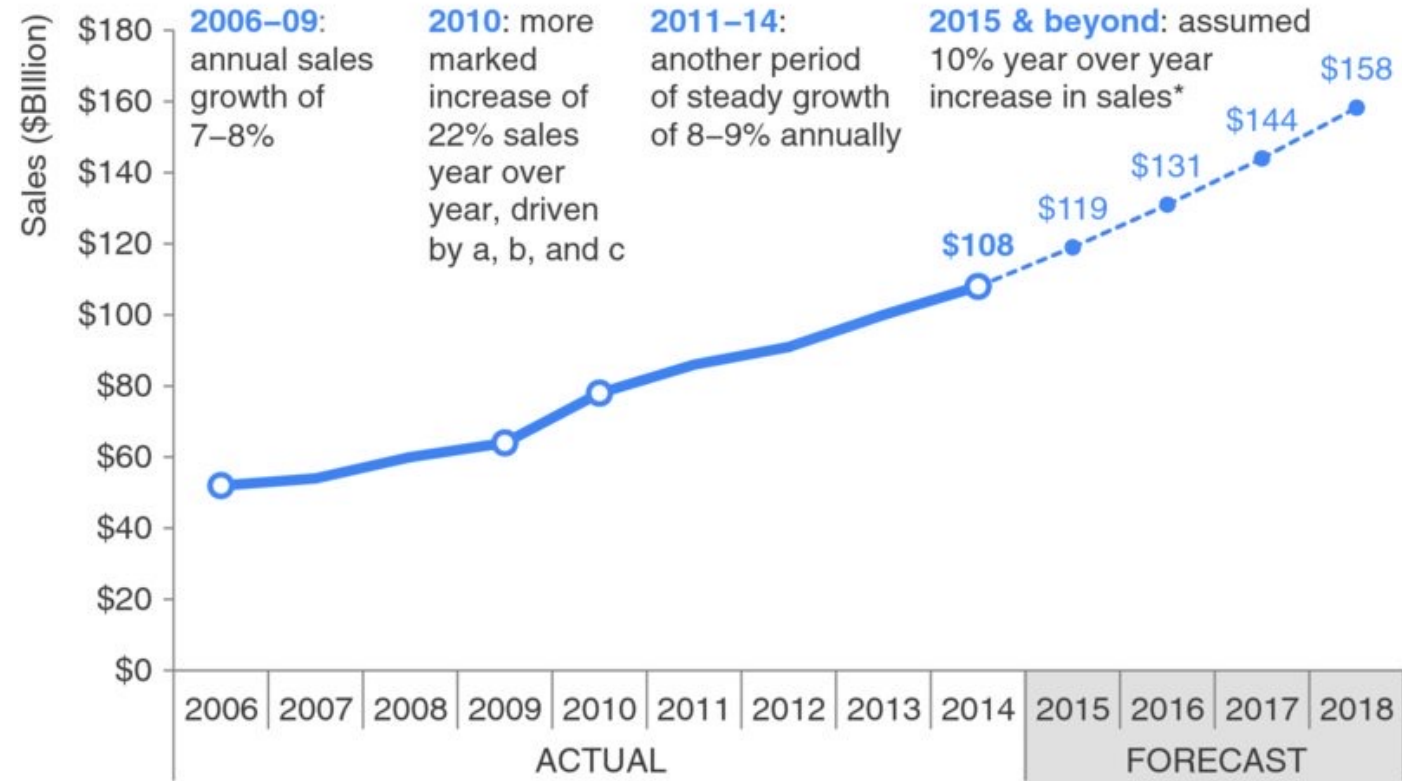
Source: Storytelling with Data

Exemple : Graphique linéaire



Exemple : Graphique linéaire et prévision

Sales over time



Data source: Sales Dashboard; annual figures are as of 12/31 of the given year.

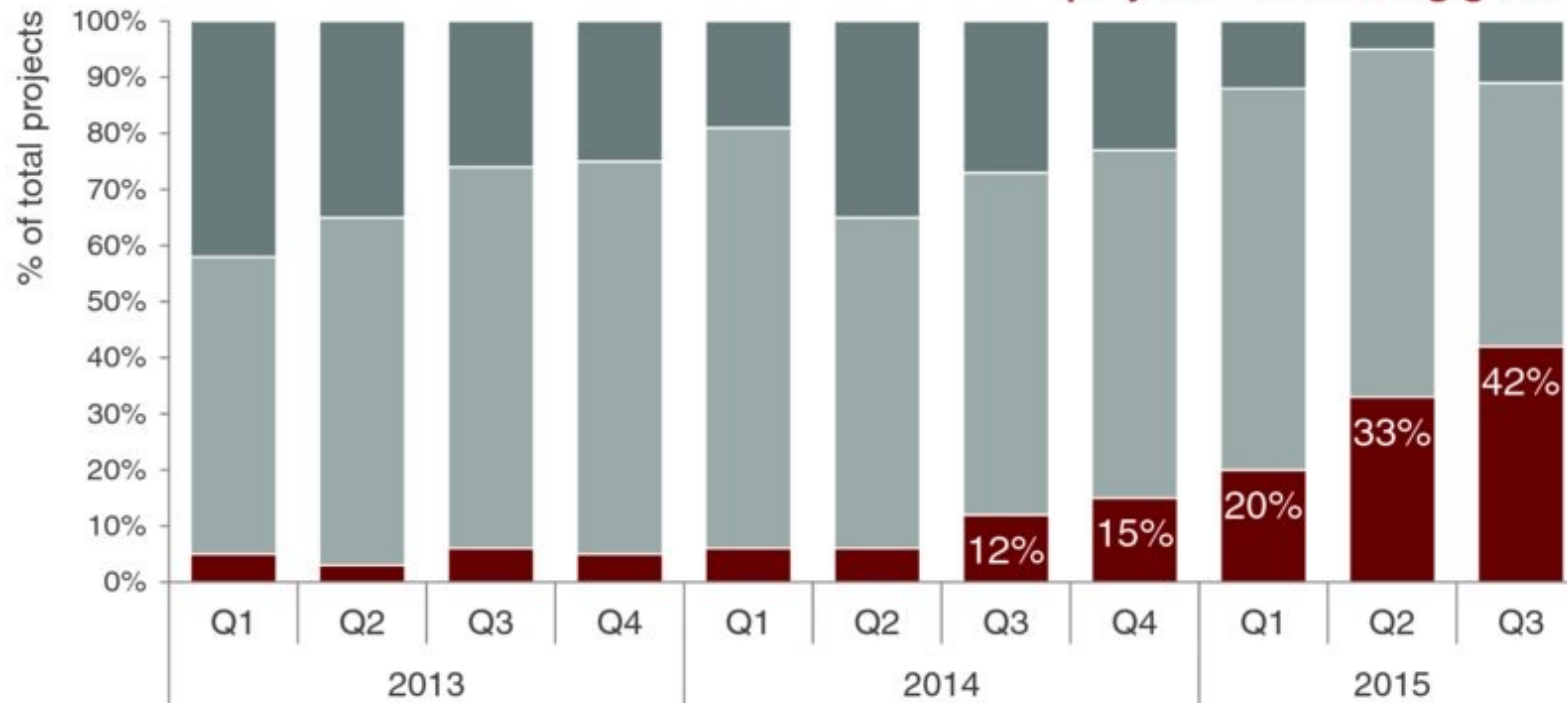
*Use this footnote to explain what is driving the 10% annual growth forecast assumption.

Exemple : Barres empilées

Goal attainment over time

■ Miss ■ Meet ■ Exceed

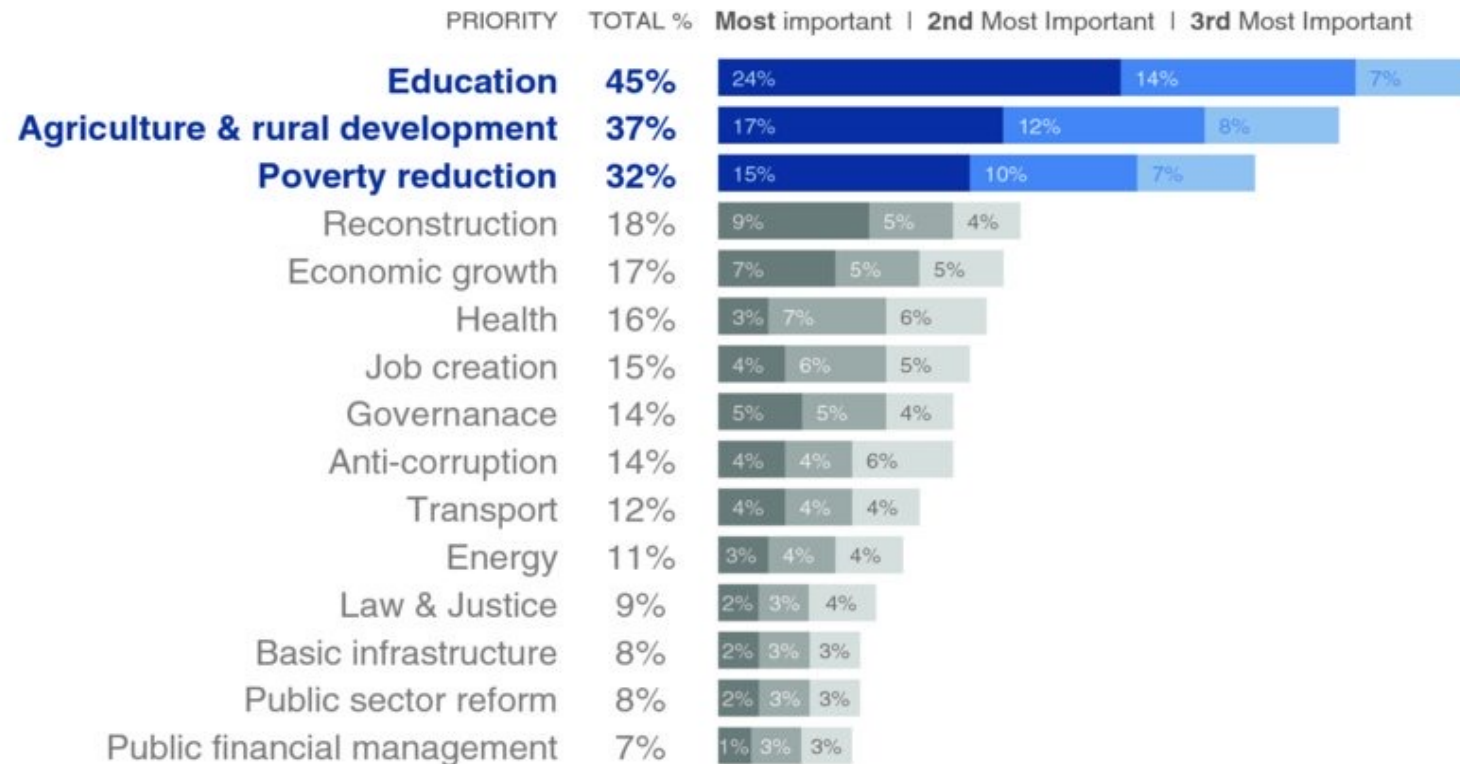
As of Q3 2015, more than 1/3 of projects are missing goals



Data source: XYZ Dashboard; the total number of projects has increased over time from 230 in early 2013 to nearly 270 in Q3 2015.

Exemple : Barres horizontales empilées

Top 15 development priorities, according to survey



N = 4,392. Based on responses to item, *When considering development priorities, which one development priority is the most important? Which one is the second most important priority? Which one is the third most important priority?* Respondents chose from a list. Top 15 shown.

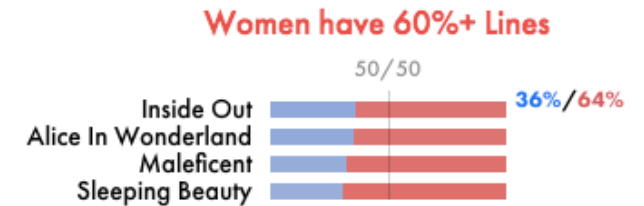
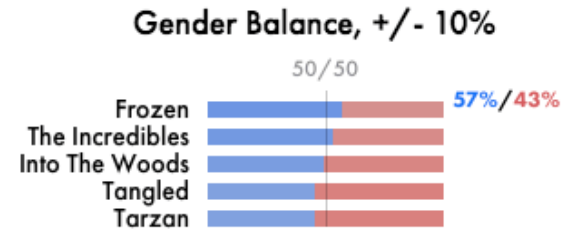
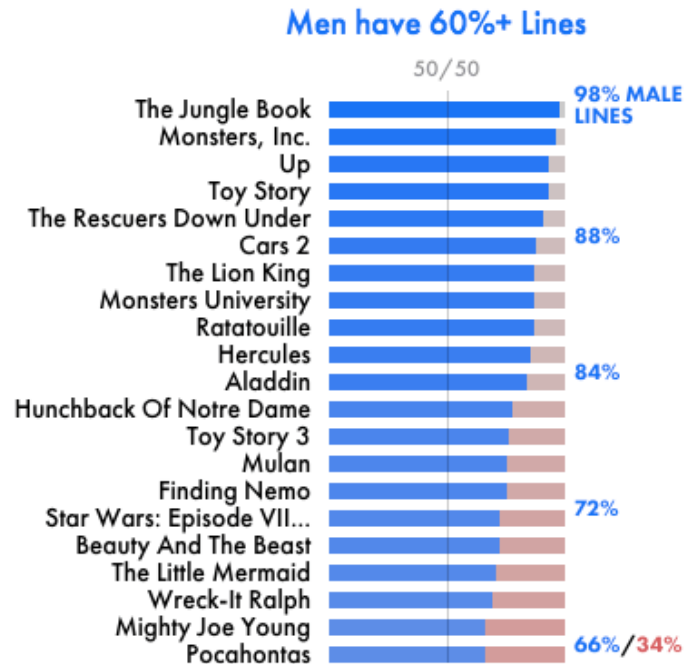
Exemple : Disney



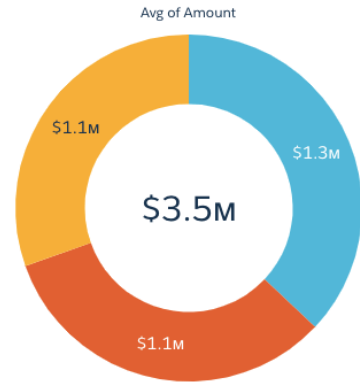
Screenplay Dialogue,
Broken-down by Gender

2,005 Screenplays: Dialogue
Broken-down by Gender

Only High-Grossing Films: Ranked in
the Top 2,500 by US Box Office*

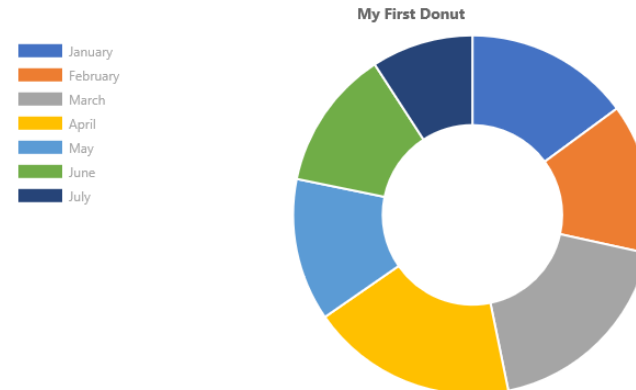
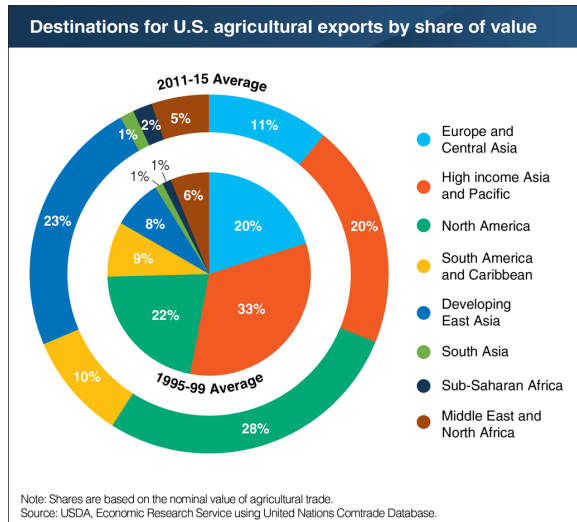
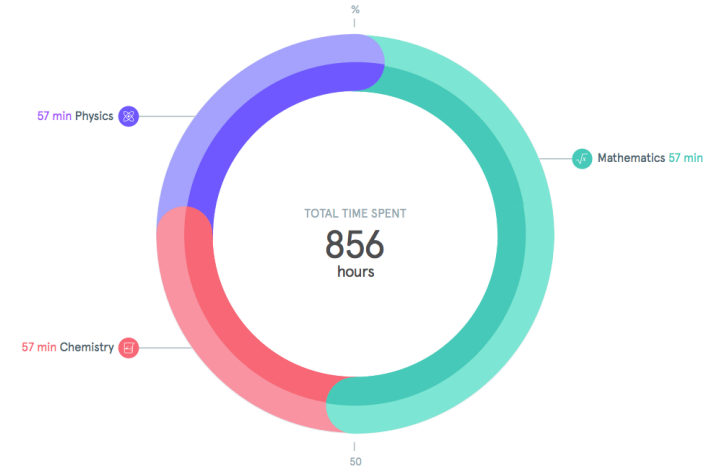


Les mauvais beignets et les bons beignets



Opportunity Type

- Existing Business
- New Business
- New Business / Add-on



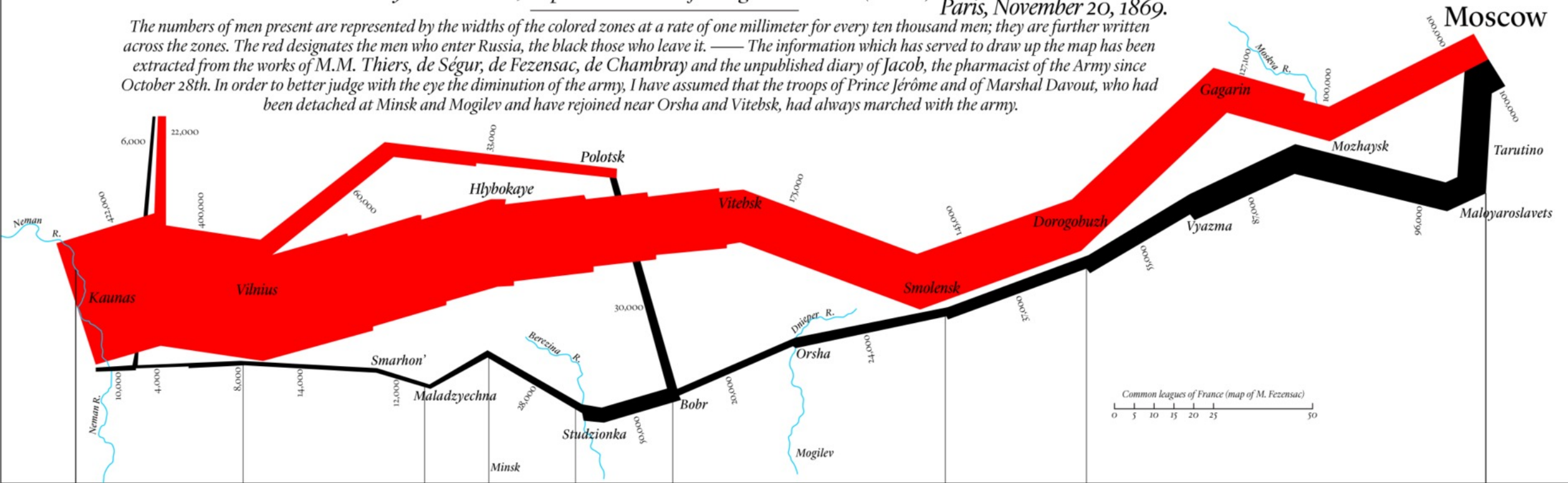
Visualisations historiques efficaces

Figurative Map of the successive losses in men of the French Army in the Russian campaign 1812 ~ 1813

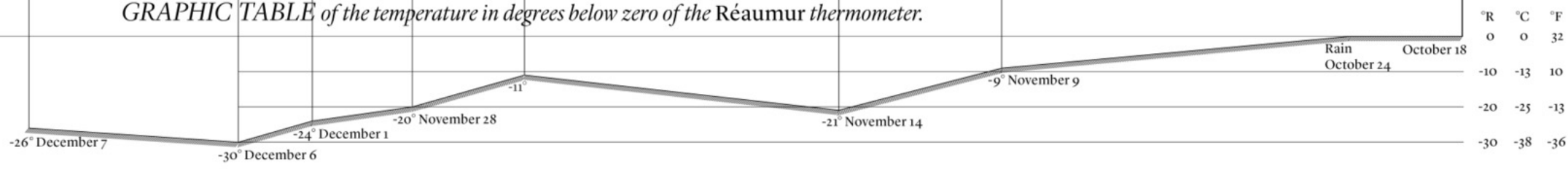
Drawn by M. Minard, Inspector General of Bridges and Roads (retired).

Paris, November 20, 1869.

The numbers of men present are represented by the widths of the colored zones at a rate of one millimeter for every ten thousand men; they are further written across the zones. The red designates the men who enter Russia, the black those who leave it. — The information which has served to draw up the map has been extracted from the works of M.M. Thiers, de Ségur, de Fezensac, de Chambray and the unpublished diary of Jacob, the pharmacist of the Army since October 28th. In order to better judge with the eye the diminution of the army, I have assumed that the troops of Prince Jérôme and of Marshal Davout, who had been detached at Minsk and Mogilev and have rejoined near Orsha and Vitebsk, had always marched with the army.



GRAPHIC TABLE of the temperature in degrees below zero of the Réaumur thermometer.



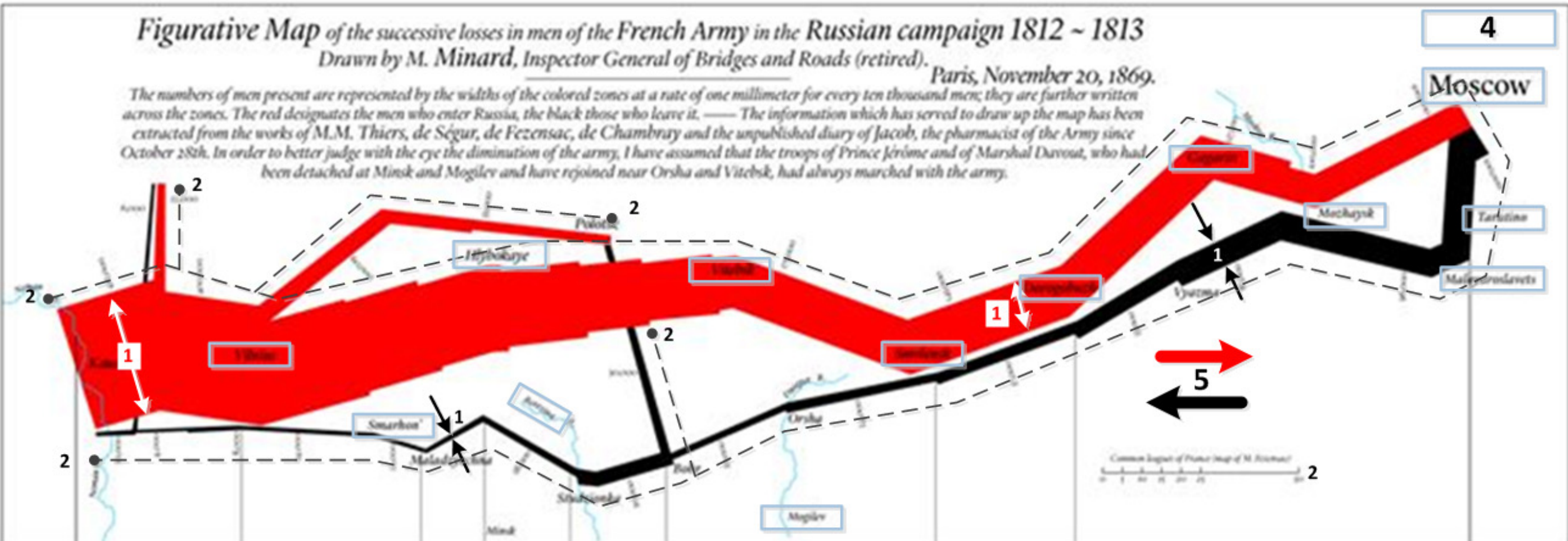
Visualisations historiques efficaces

Figurative Map of the successive losses in men of the French Army in the Russian campaign 1812 ~ 1813

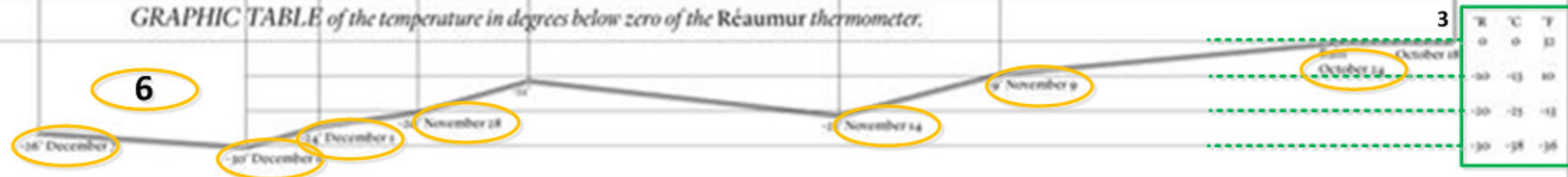
Drawn by M. Minard, Inspector General of Bridges and Roads (retired).

Paris, November 20, 1869.

The numbers of men present are represented by the widths of the colored zones at a rate of one millimeter for every ten thousand men; they are further written across the zones. The red designates the men who enter Russia, the black those who leave it. — The information which has served to draw up the map has been extracted from the works of M.M. Thiers, de Ségur, de Fozensac, de Chambray and the unpublished diary of Jacob, the pharmacist of the Army since October 28th. In order to better judge with the eye the diminution of the army, I have assumed that the troops of Prince Jérôme and of Marshal Davout, who had been detached at Minsk and Mogilev and have rejoined near Orsha and Vitebsk, had always marched with the army.

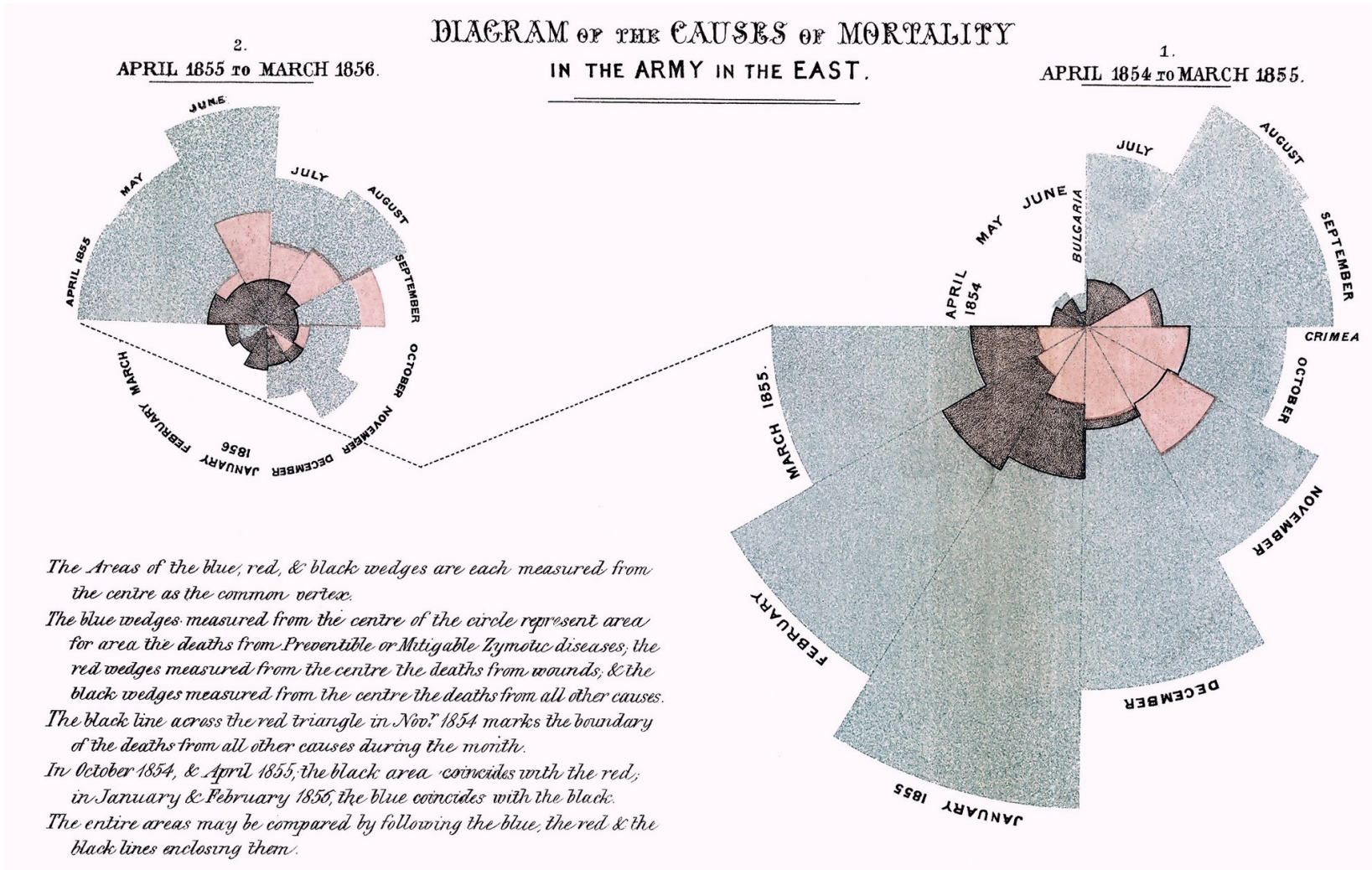


GRAPHIC TABLE of the temperature in degrees below zero of the Réaumur thermometer.



The Cossacks pass the frozen Nieman at a gallop.

Visualisations historiques efficaces



Pendant la Seconde Guerre mondiale, le mathématicien A. Wald a entrepris une étude pour aider à protéger les bombardiers britanniques qui survolaient le territoire ennemi.



Les données comprenaient : **le nombre** et **l'emplacement** des impacts de balles sur les avions retournés, et l'objectif était d'utiliser ces informations pour déterminer où ajouter un blindage pour protéger au mieux la structure de l'avion.

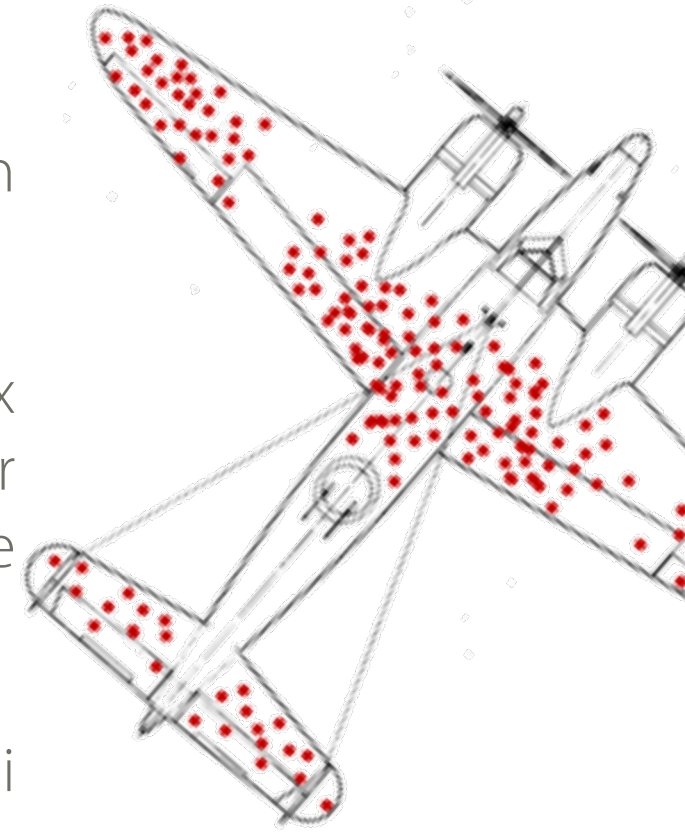
Un graphique a été créé pour montrer où le nombre maximum d'impacts de balles était situé sur les **avions retournés**. Ce graphique montrait que les dommages les plus importants se situaient aux **extrémités** de l'avion, et non sur les longerons principaux de l'aile et de la queue, les moteurs et les parties centrales du fuselage.

L'histoire de Wald

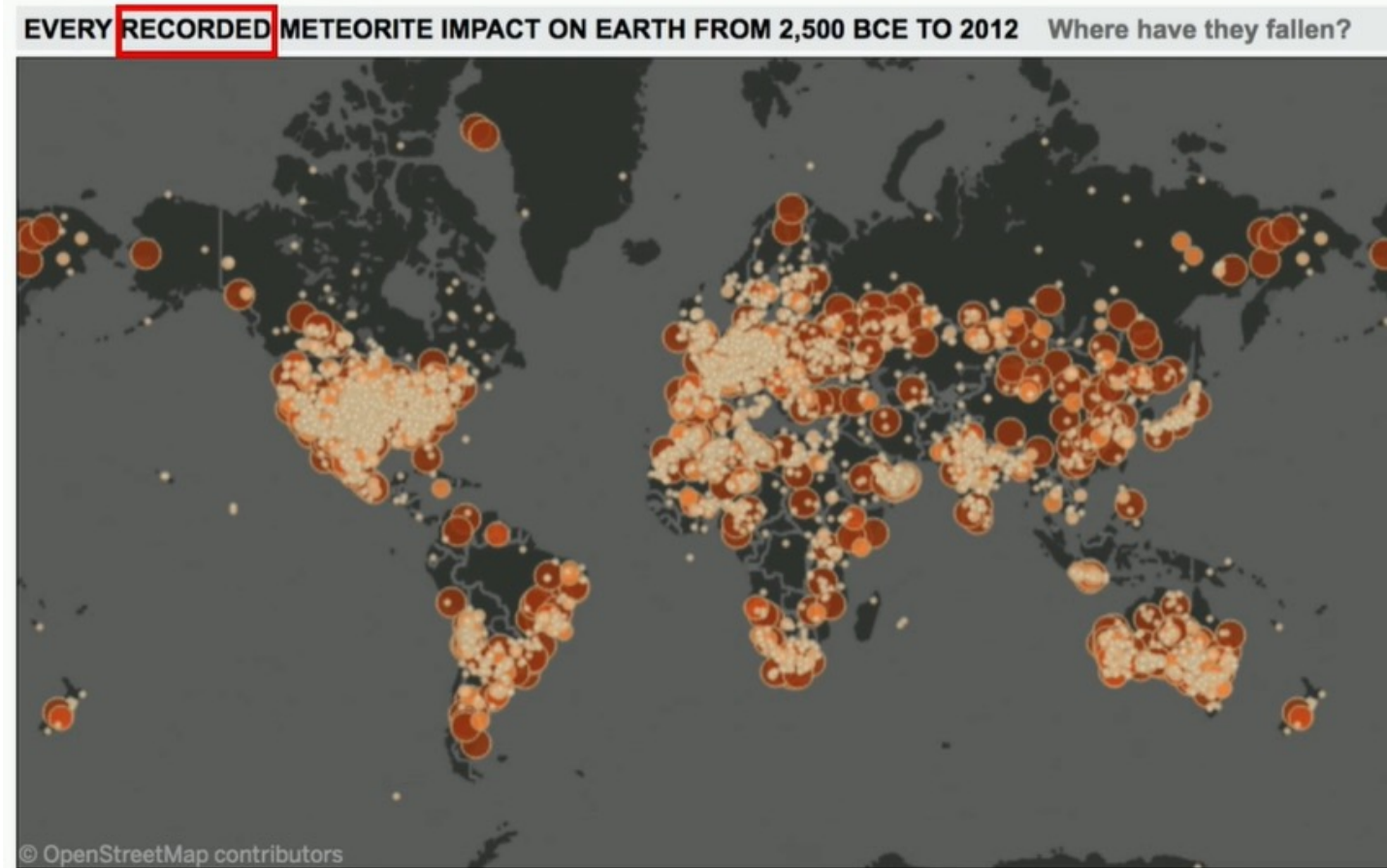
Ainsi, le ministère de la défense aérienne voulait ajouter un blindage aux **extrémités**. Wald a suggéré qu'ils avaient **tout faux**.

Pour éviter le "**biais de survie**", le blindage doit être ajouté aux zones présentant le **moins de trous** : si aucun avion de retour n'avait de trous dans ses longerons d'aile et ses moteurs, même quelques trous à ces endroits étaient **mortels**.

À retenir : les données manquantes peuvent être aussi importantes pour le récit que les données présentes. Raconter une histoire n'est pas toujours une entreprise évidente.

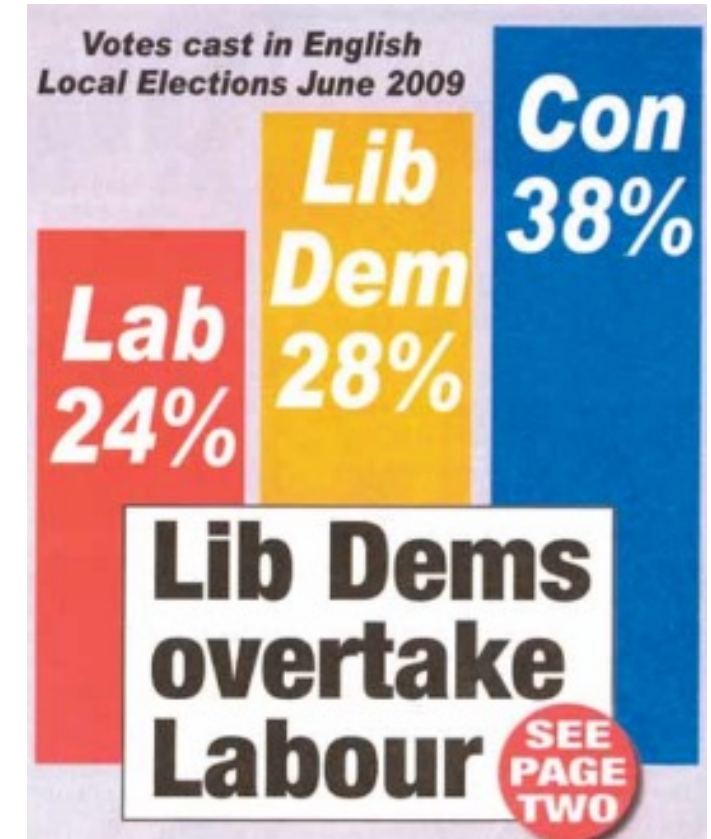


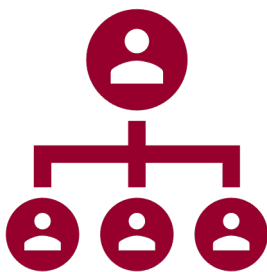
Prendre ce qui est mesuré pour la vérité



Source: The Head Game

Graphiques trompeurs





Module 6

L'ingénierie et la gouvernance des données

Démocratisation de l'analyse des données

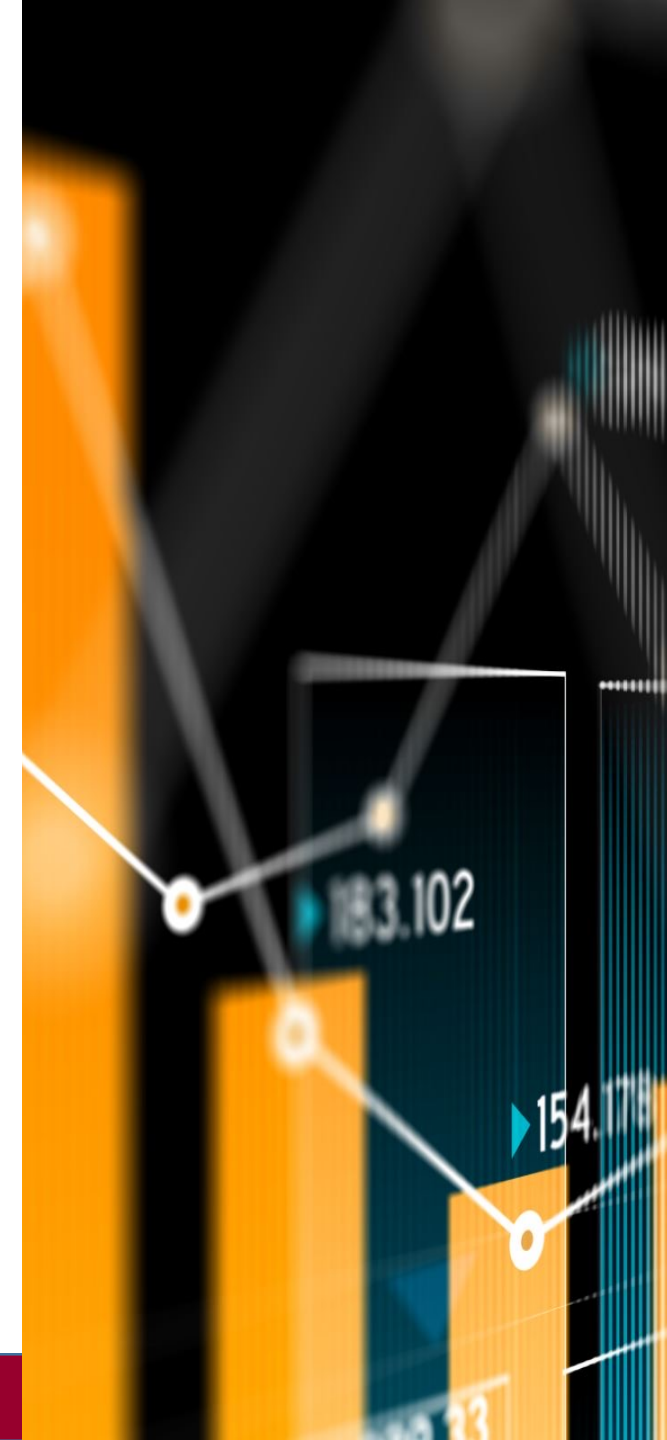
Les besoins en intelligence économique poussent au développement d'outils d'analyse de données et de pipelines pour les non experts en données, comme par exemple :

- PowerBI ou Tableau

Démocratisation des données + augmentation des connaissances en matière de données/numérique.

Cela va probablement pousser les organisations à aller de l'avant également.

Ne remplace pas **nécessairement** les pipelines de données "industriels" ou "professionnels".



**De la BI
(*business
intelligence*) à
l'AI/ML
(*apprentissage
automatique*)**

Dans une certaine mesure, la mise en place et le fonctionnement d'un solide pipeline de BI professionnelle/industrielle est une étape importante dans une organisation.

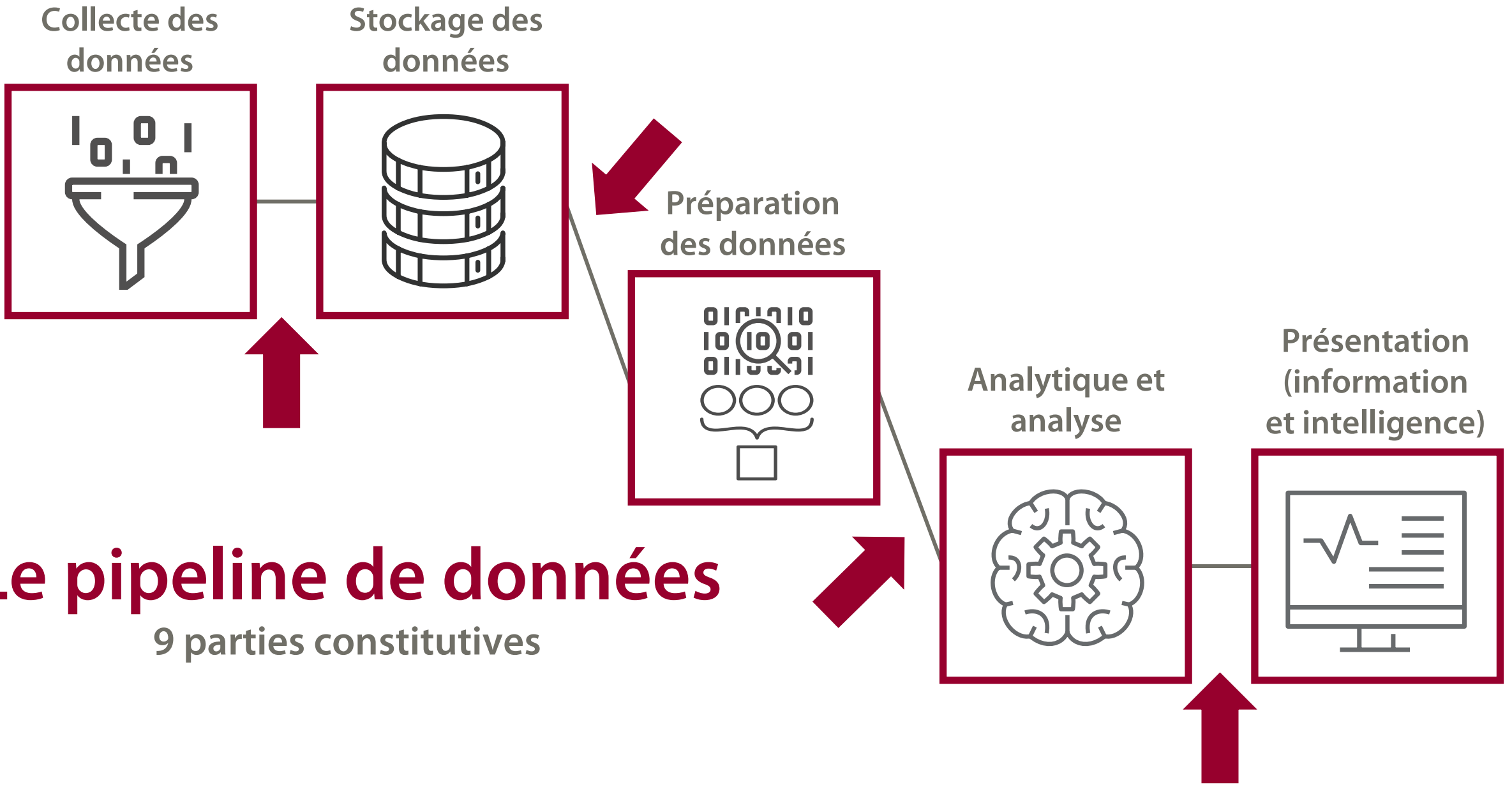
MAIS - l'architecture de données et les outils dont vous avez besoin pour l'analyse AI/ML/DS peuvent ne pas être les mêmes que ceux de la BI.

Vous devrez PEUT-ÊTRE revoir la conception de certaines parties de votre pipeline BI pour prendre en charge l'IA/ML/DS.

En particulier, l'architecture de votre base de données : Data Lakes vs DataMart vs NoSQL

Points à prendre en compte lors du choix des outils d'analyse

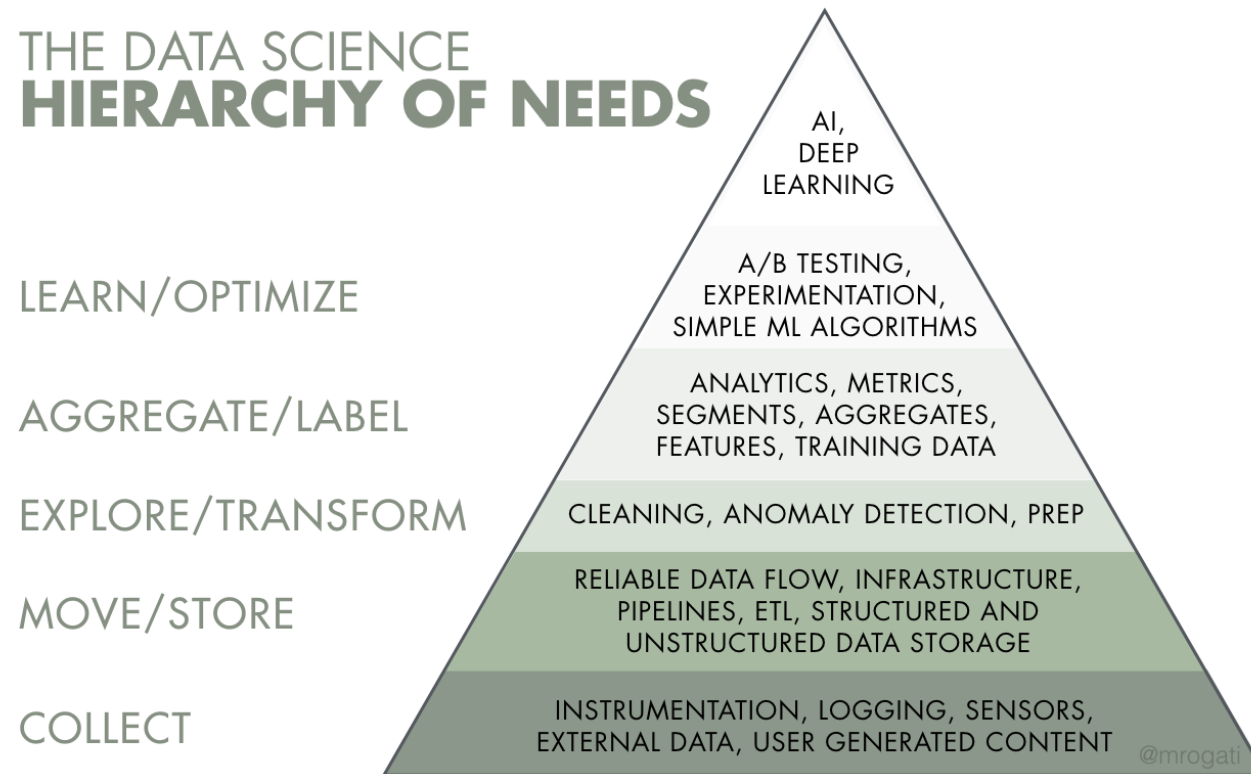
- A. **Capacité** : quelles sont leurs fonctionnalités et leurs performances - disposent-elles de toutes les techniques, de la puissance de traitement ?
- B. **Intégration** : comment se connectent-ils aux autres parties de votre pipeline ?
- C. **Expérience de l'utilisateur** : quelle est l'expérience de l'utilisateur - quel est le niveau d'expertise nécessaire pour utiliser cet outil, quelle est la facilité d'utilisation de cet outil ?
- D. **Coût** : court et long terme



Le pipeline de données

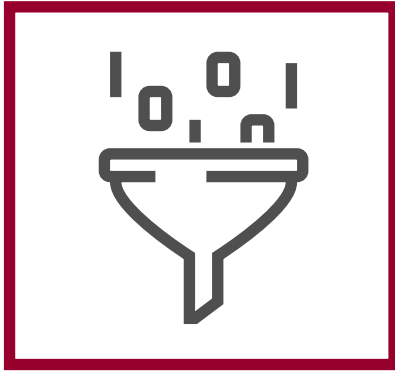
9 parties constitutives

Hiérarchie des besoins en matière de données

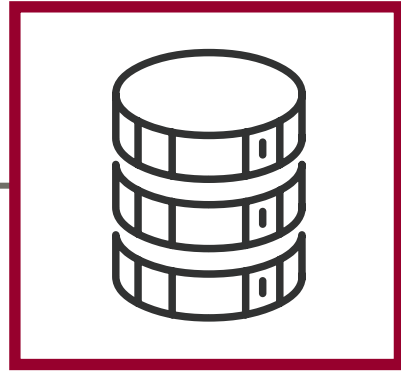


Pipeline de données à source ouverte

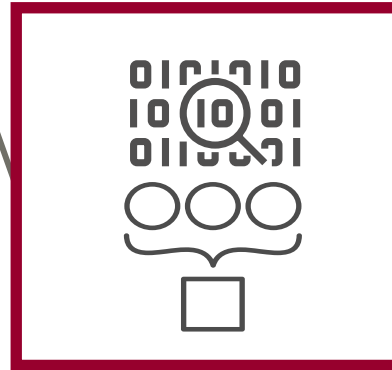
Collecte des
données



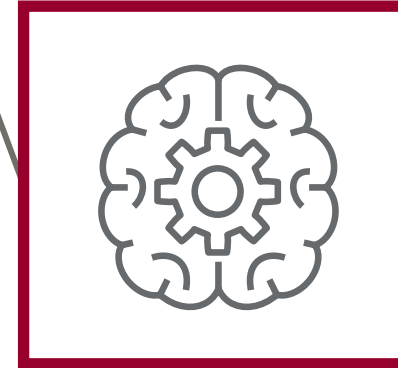
Stockage des
données



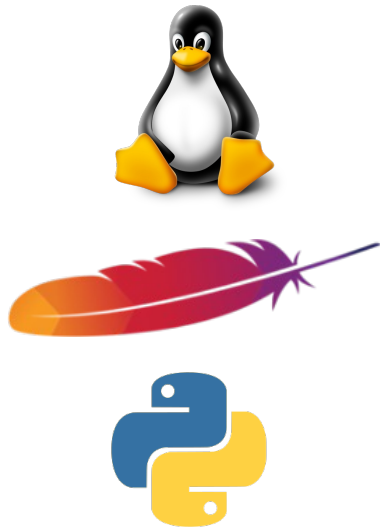
Préparation
des données



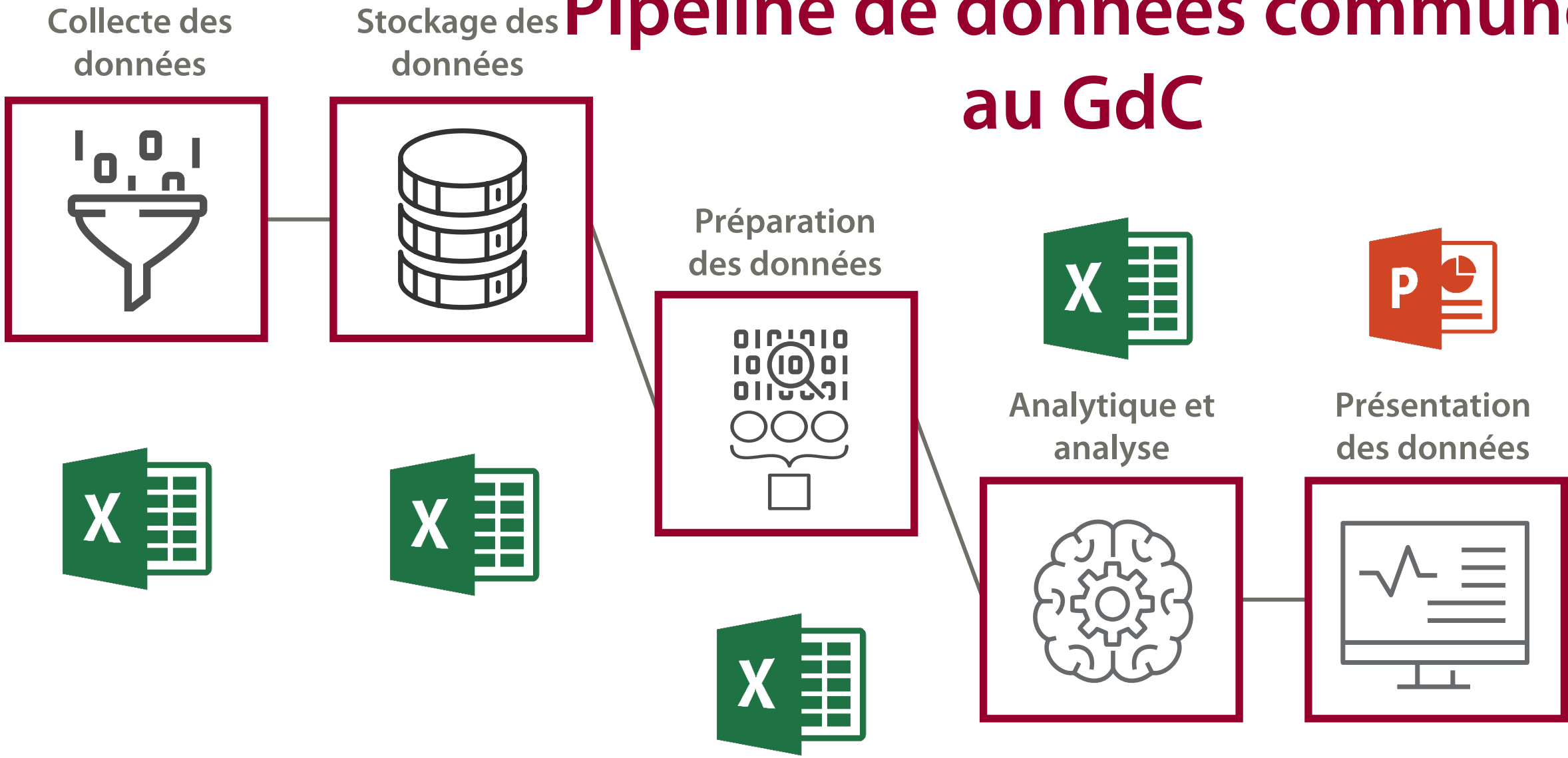
Analytique et
analyse



Présentation
des données



Pipeline de données commune au GdC



Technologies de pipeline de données

	Sur site (LAN)	Cloud public	Cloud privé
Amateur	Répertoire partagé + Excel + Power Point + Accès au bureau	Technologie de pipeline automatisé de bout en bout basée sur un serveur : Sur site Azure, sur site IBM RedHat	Solutions maison utilisant des serveurs installés dans le nuage - par exemple, AWS, GCP.
Semi-professionnel	'Desktop' Science des données : Desktop PowerBI SQL-Lite (Desktop) MS Access Autonome interne DBMS – lecture et écriture	Pipeline de données SaaS de bout en bout : par exemple, COTS Pachyderm ou plus personnalisé : par exemple, SaaSCoder.	Données en nuage de bout en bout Pipeline Infrastructure(sans serveur/ sans serveur) : AWS, GCP, Azure
Professionnel	Technologie de pipeline automatisé de bout en bout sur serveur : Azure sur site, IBM RedHat sur site		

Comprendre le contexte de l'informatique dématérialisée



IaaS: (Infrastructure as a Service)
L'infrastructure en tant que service



PaaS: (Platform as a Service)
Plate-forme en tant que service

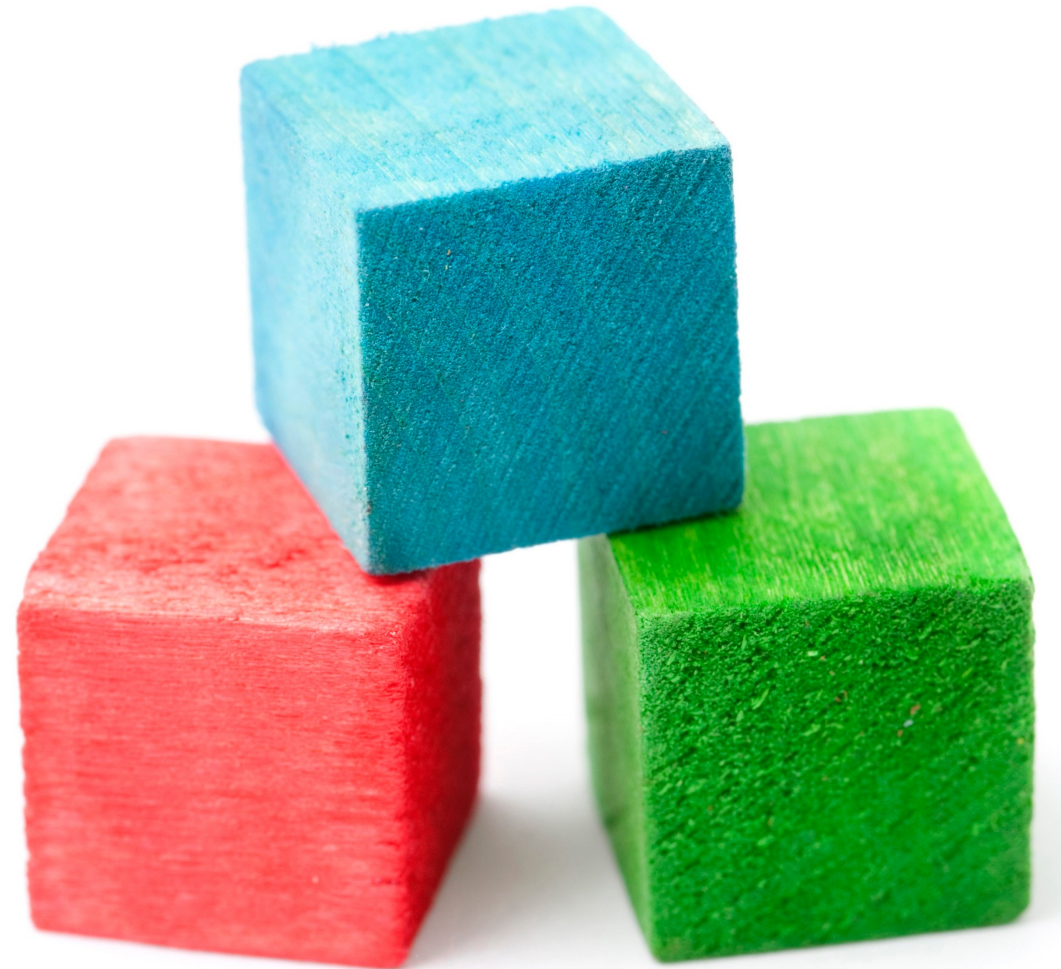


SaaS (AlaaS, DaaS):
(Software (AI, Data) as a Service)
Logiciels (IA, données) en tant que service

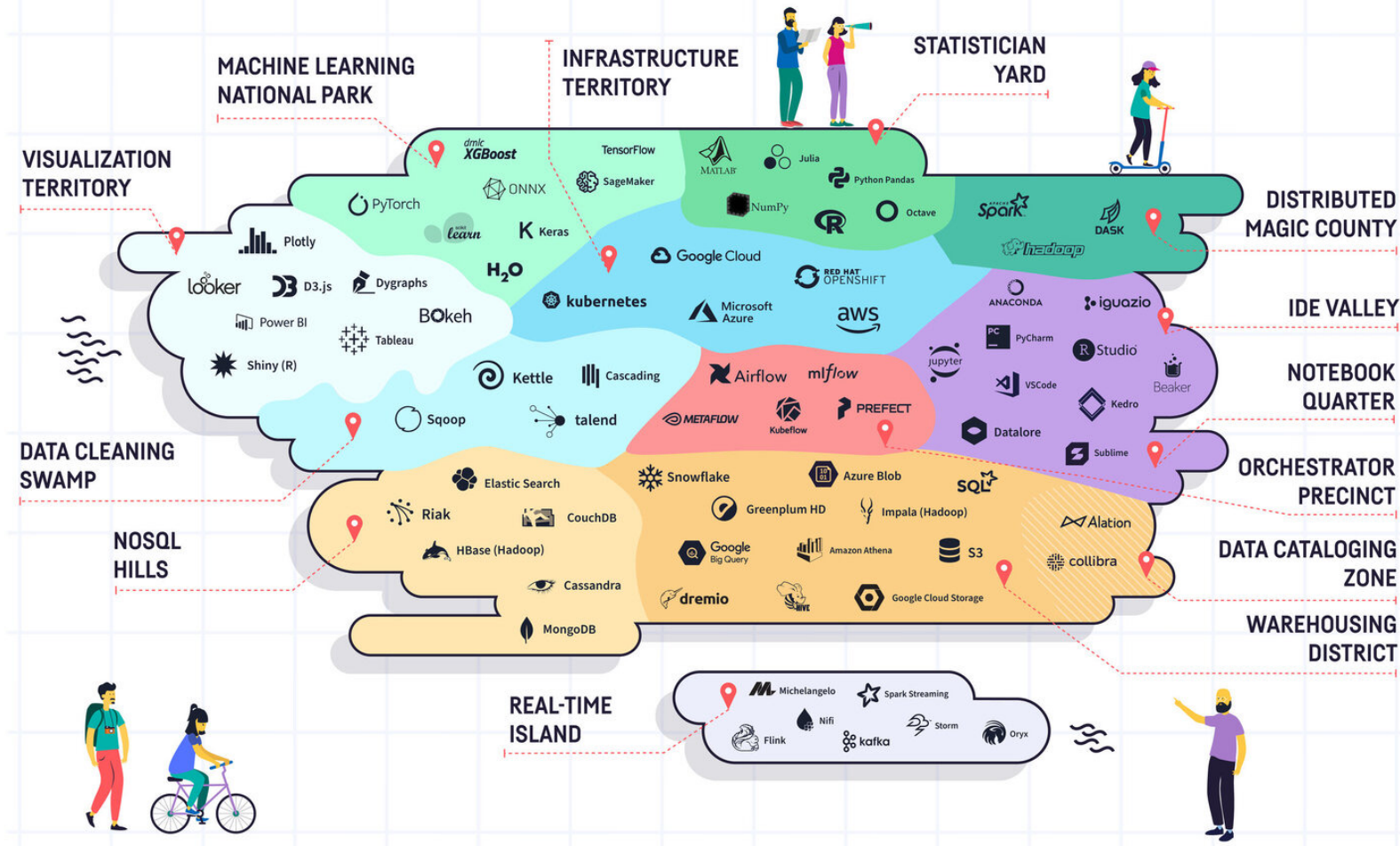
Phases de création des pipelines

1. Recherche et conception
2. Mise en œuvre
3. Essais
4. Production + gestion
5. Recherche et conception

Agile !



THE LONG JOURNEY ACROSS
Technoslavia



Construire
ou acheter ?

**What does the future of
the data world look like?**

An empire in which a single language dominates all others?
Or a unified world where multiple republics of data technologies coexist and complement each other?
At Dataiku, we believe in the latter. And in this world (which we've dubbed Technoslavia), data teams need a reassuring guide; a tool that is always on the forefront of the latest technologies and that also unites all profiles, from data scientist to analyst to IT.



Construire plutôt qu'acheter

1. Vous voulez obtenir un avantage compétitif.
2. Besoin de rattraper les concurrents/autres départements.
3. Tenter d'encourager un plus grand engagement des clients ou des employés.
4. Création potentielle d'une nouvelle source de revenus.

Certaines personnes construisent puis, une fois construites, se rendent compte que leur solution est mauvaise.



Qui prend cette décision en général ?

C-Suite, responsable des données, chef de produit, chef de projet

- idéalement, un conseiller technique ?

Conséquences de la **construction** en interne :

- les normes de l'industrie changent
- problèmes de maintenance
- dépendance à l'égard du talent
- la création de connaissances tacites

Chronologie de la décision d'achat

Si vous regardez les choses du seul point de vue des coûts, vous pouvez décider rapidement.

Vous pourriez avoir besoin de :

- une personne chargée de l'application
- responsable de l'infrastructure
- personne chargée de l'architecture des données
- Responsable UI/UX

Biais de construction : les techniciens de l'entreprise peuvent être enthousiasmés par l'idée de construire en raison de leur expérience, parce qu'ils sont passionnés.

Construire vs acheter

Faites preuve de diligence raisonnable ; il est facile d'être partial à l'égard d'un fournisseur.

Obtenez le soutien nécessaire pour commencer l'enquête.

Ne vous fiez pas uniquement au nombre de fonctionnalités :

- Combien de clients ont eu du succès avec le fournisseur ?

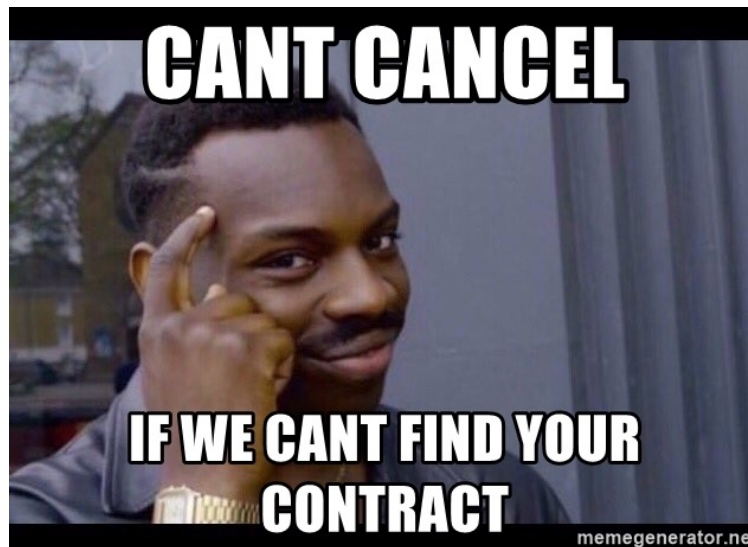
Trouver le bon partenaire :

- Développement de logiciels personnalisés
- Forte expertise en matière de sécurité et d'infrastructure
- Bonne intégration dans la gouvernance des données existante
- Un partenaire sur lequel vous pouvez vous appuyer

Conséquences d'une mauvaise décision

Dette technique de l'internalisation et de l'externalisation.

Si vous êtes coincé avec le mauvais fournisseur, vous pouvez vous retrouver dans le pétrin sans le savoir.



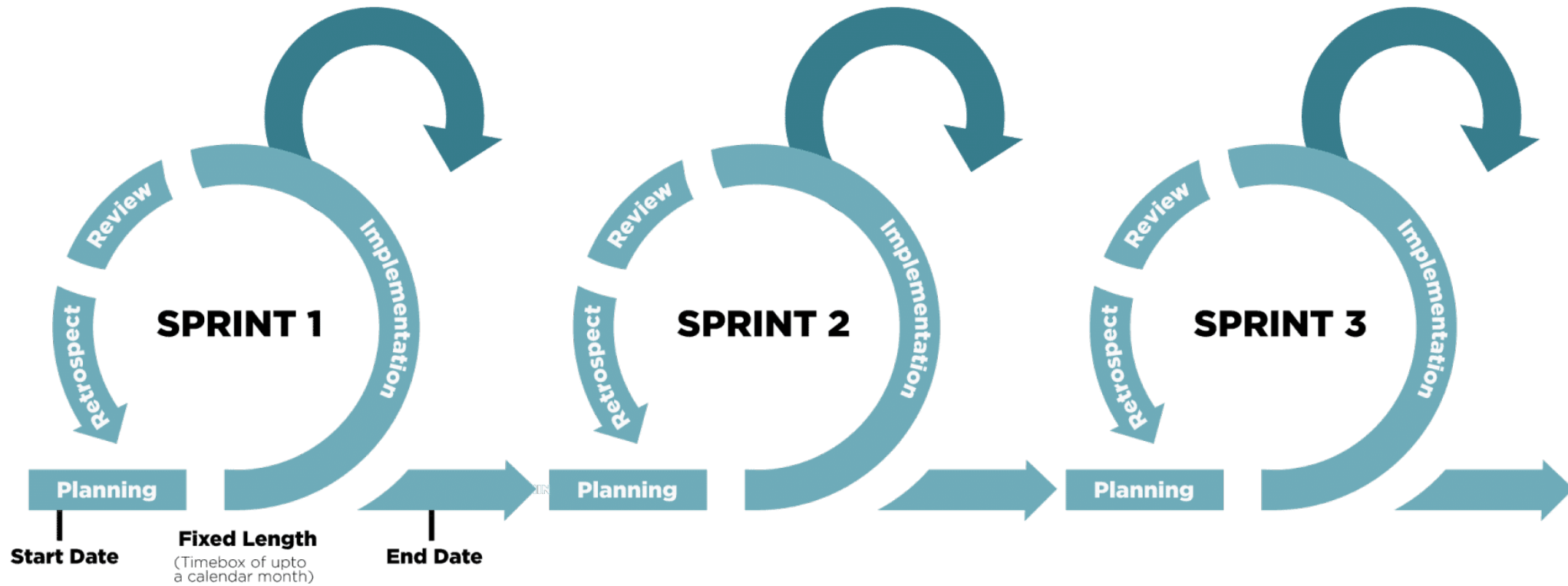
Drapeaux rouges (« red flags ») lors de la sélection des fournisseurs

Ils promettent des fonctionnalités qui seront livrées après l'achat :

- Livraison incohérente de la feuille de route.
- Les exigences architecturales intégrées sont coûteuses.
- Impossibilité d'utiliser le logiciel sans recourir à des services professionnels.



Agile



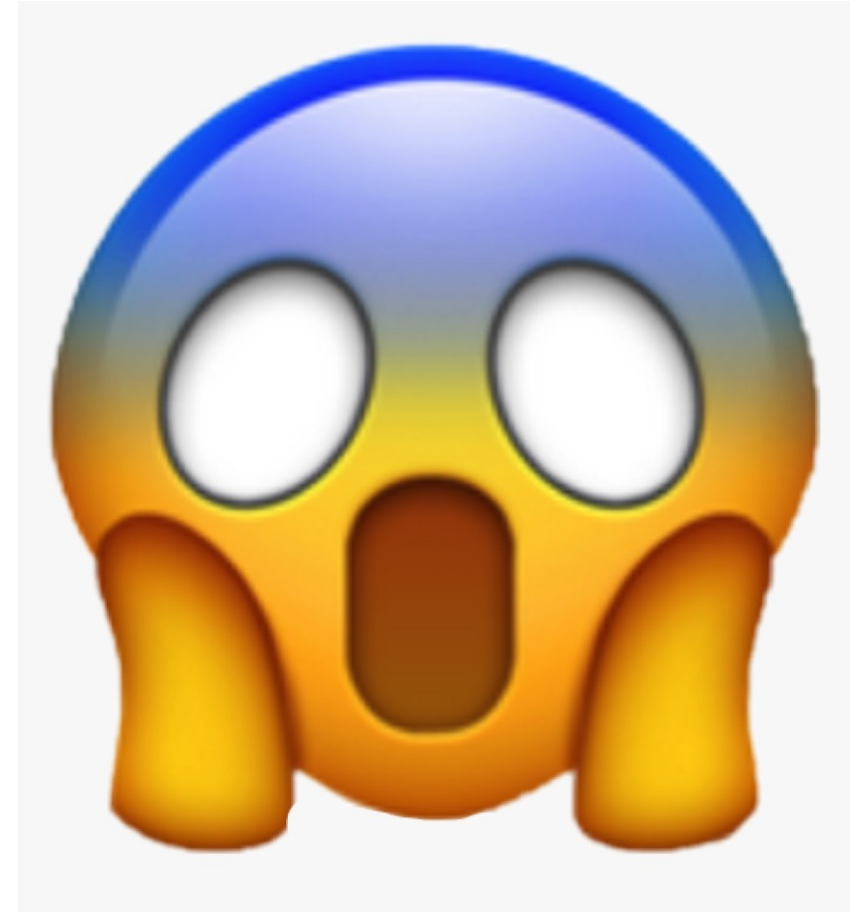
La malédiction des données catégorielles

Les données du GdC ont tendance à être très lourdes en ce qui concerne les catégories et les données textuelles.

Méthodes d'analyse traditionnelles :

- ne met pas l'accent sur les données catégorielles.
- ne s'est pas concentré sur la réalisation d'analyses complexes avec des données catégorielles (complexes).

Cela signifie que nous devons travailler plus dur pour produire de bonnes stratégies pour traiter ce type de données (indice : ML aime les catégories).

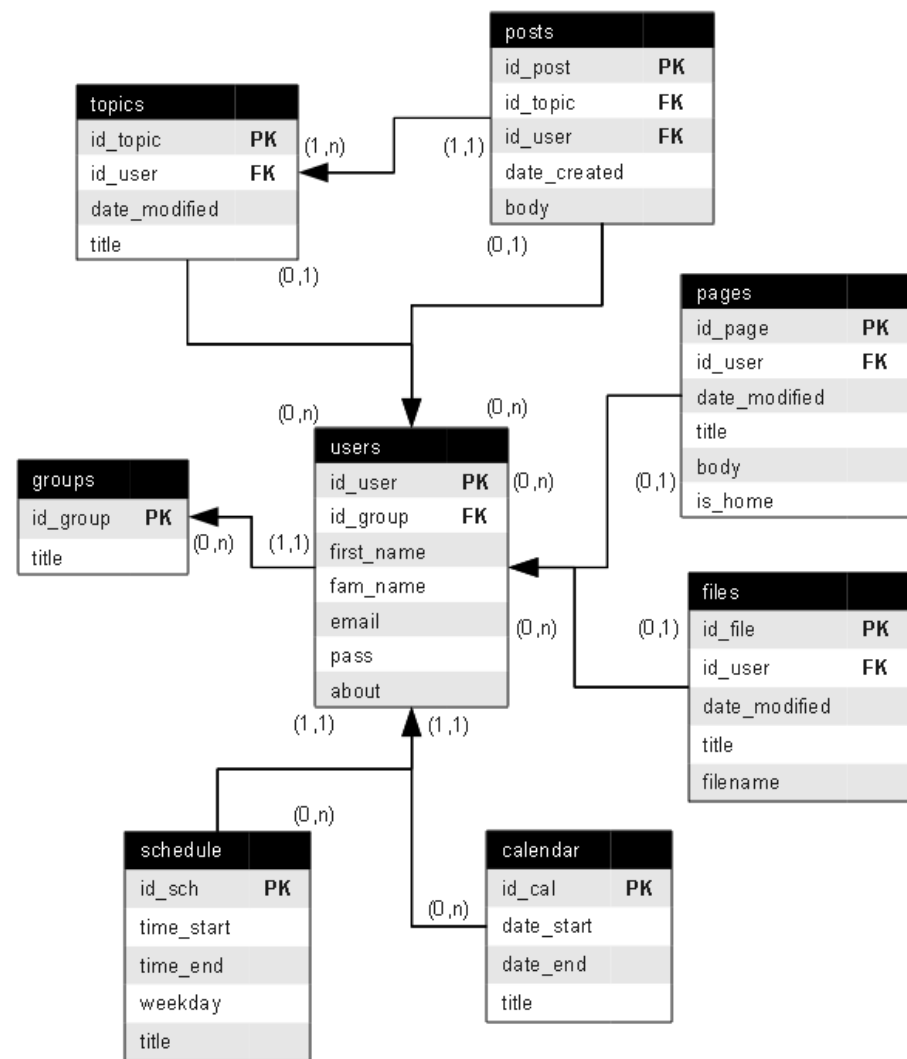


Stockage des données

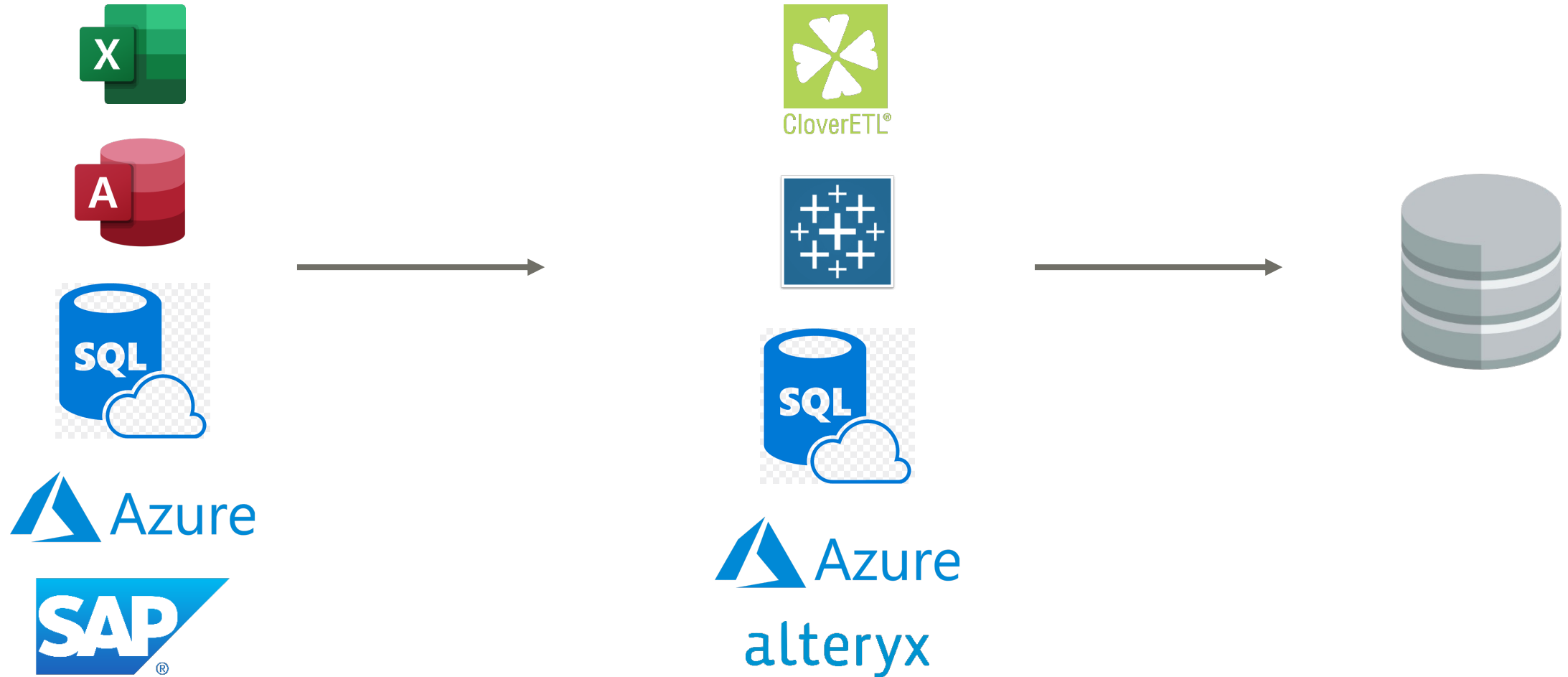
Quatre options différentes qui sont actuellement populaires en termes de données fondamentales et de stratégies de modélisation ou de structuration des connaissances :

- des paires clé-valeur (par exemple JSON)
- triples (par exemple, le cadre de description des ressources [RDF])
- bases de données graphiques
- bases de données relationnelles

C'est plus important qu'il ne l'apparaît...



Comment nettoyer les données : Perspective de l'outillage



Qu'est-ce que la gouvernance des données ?

La gouvernance des données est un concept qui permet à une organisation de garantir la qualité des données tout au long de leur cycle de vie.

En se concentrant sur la gouvernance des données, on s'assure que les données de l'ensemble de l'organisation :

- sont **disponibles** quand on en a besoin
- sont **utilisables** lorsqu'on y accède
- sont **cohérentes** lorsqu'elles sont analysées
- sont **intègres** et de **haute qualité**
- sont **sécurisées** et **dignes de confiance**

Qu'est-ce que la gouvernance des données ?

La gouvernance des données comprend :

- **les personnes, les processus, et les technologies de l'information**

Elle est nécessaire pour assurer un traitement cohérent et approprié des données d'une organisation à travers l'entreprise.

Elle fournit la base, la stratégie et la structure nécessaires pour garantir que les données sont gérées comme un actif et transformées en informations significatives.

Exercice : en équipe de deux, énumérez les cas où vous avez rencontré des problèmes liés à la disponibilité, à la fonctionnalité, à la cohérence, à l'intégrité, à la qualité, à la sécurité ou à la fiabilité des données !

La gouvernance des données au sein du GdC



Point de référence central pour le GdC :

- [Plans stratégiques, politiques, normes et lignes directrices relatives aux services numériques du gouvernement](#)

Rapport au greffier du Conseil privé :

- [Feuille de route de la Stratégie de données pour la fonction publique fédérale - Bureau du Conseil privé](#)

Secrétariat du Conseil du Trésor (sélectionné) :

- [Politique sur les services et le numérique](#)
- [Plan stratégique du gouvernement du Canada pour la gestion de l'information et la technologie de l'information de 2017 à 2021](#)
- [Plan stratégique des opérations numériques de 2018 à 2022](#)
- [Stratégie d'adoption de l'informatique en nuage du gouvernement du Canada : Mise à jour de 2018](#)

Innovation, Sciences et Développement économique Canada :

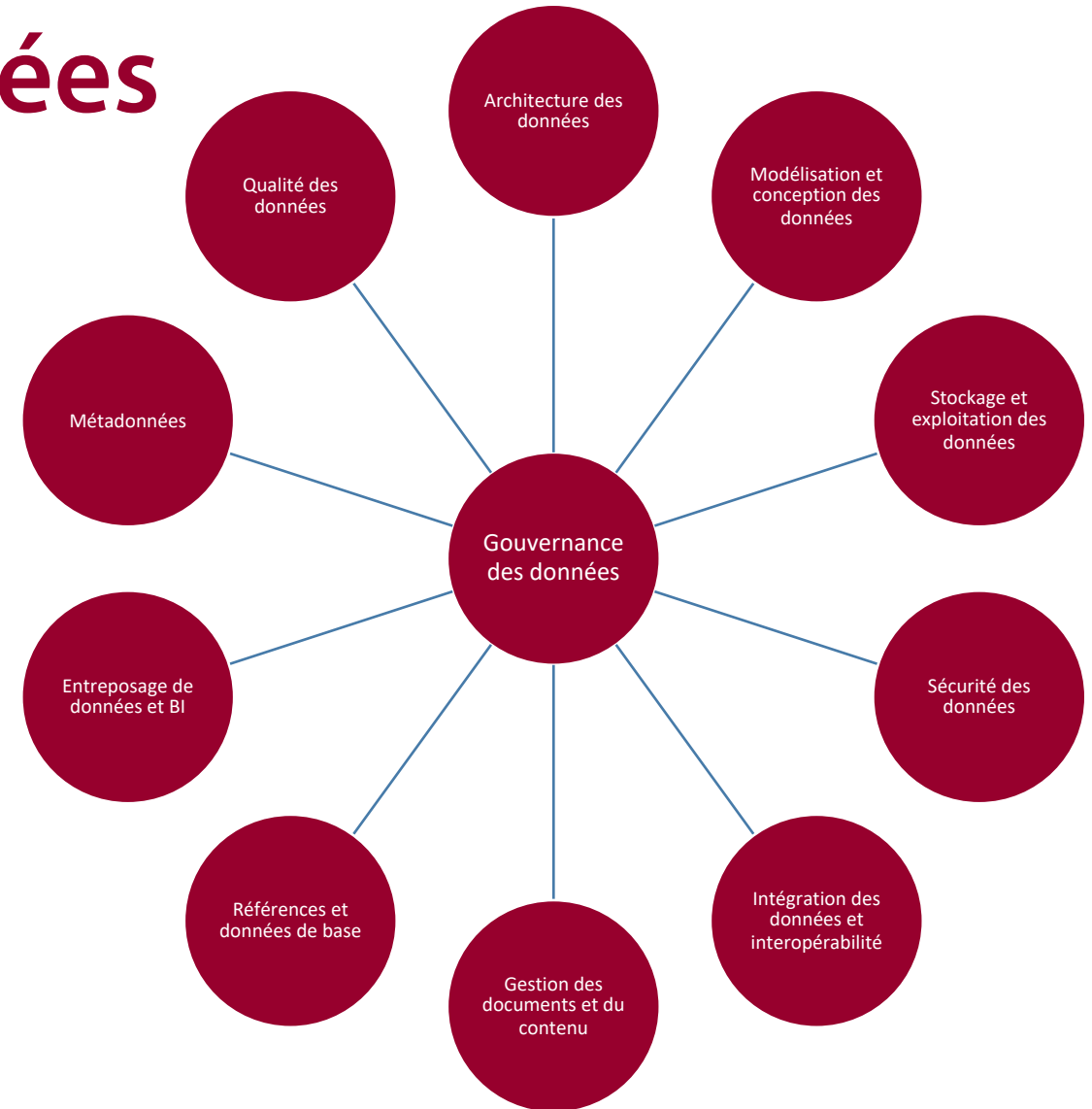
- [La Charte numérique du Canada en action: un plan par des Canadiens, pour les Canadiens](#)

Gouvernance des données

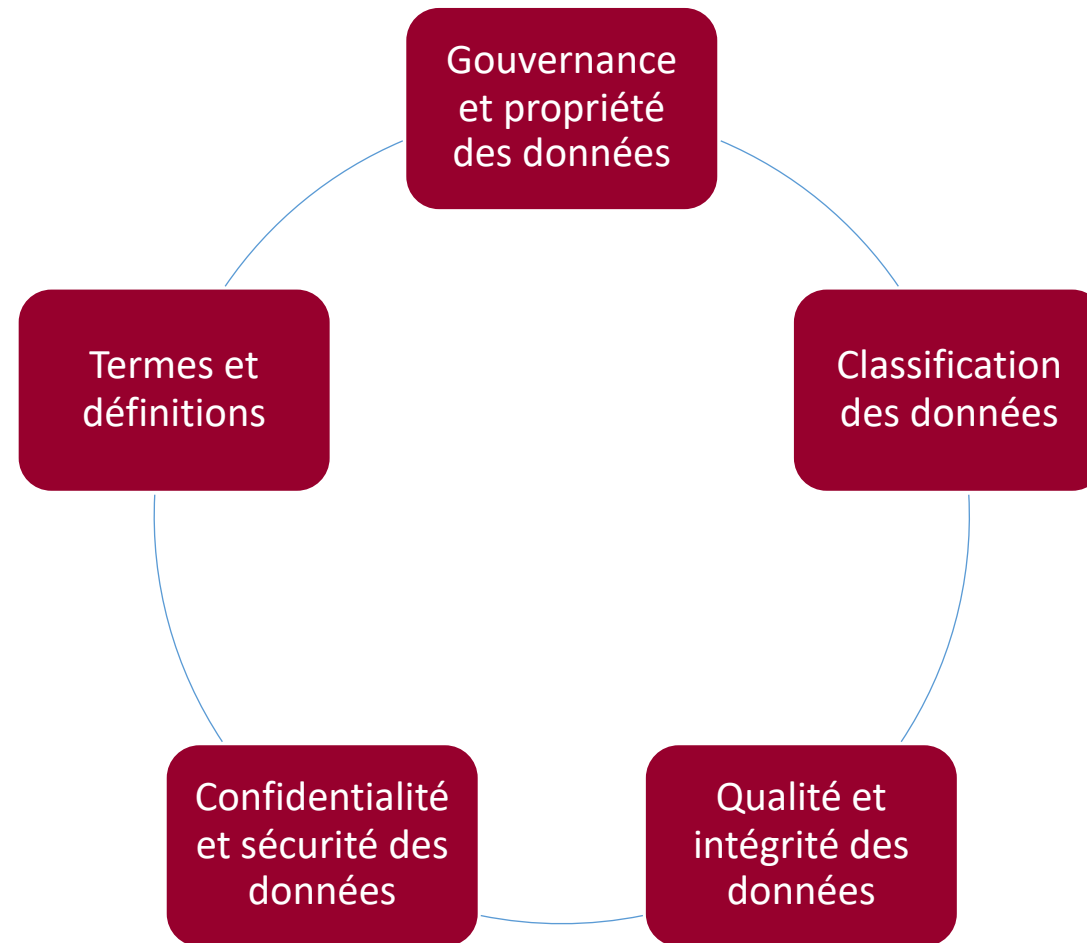
Association pour la gestion des données
(DAMA)

DMBOK 2 (Corpus de connaissances en
gestion des données) :

- Bien détaillé et réfléchi
- Les sections sont raisonnablement alignées sur l'approche du gouvernement du Canada
- Soutenu par les organisations professionnelles
- Non axé sur le gouvernement



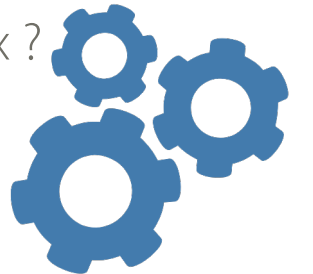
Gouvernance des données



Alignement des besoins en données

Exercice : en groupes, répondez aux questions suivantes

- votre groupe crée-t-il ou génère-t-il des données ? si oui, quelles données ?
- utilisez-vous des données provenant de sources extérieures à votre groupe ? si oui, lesquelles ?
- combien de sources de données (par exemple, des bases de données) votre groupe utilise-t-il, en général ?
- Publiez-vous l'analyse des données en interne à votre groupe ? en externe ? les deux ?



Obligations : compte-rendu et responsabilité

	Propriétaire des données	Responsable des données	Dépositaire des données	Contributeur de données	Consommateur de données
Définition	Une personne qui est responsable de la gouvernance et de la mise en conformité d'un ensemble de données.	Une personne qui est chargée de la gestion d'un ensemble de données.	Une personne qui a la responsabilité technique d'un ensemble de données.	Une personne qui crée ou collecte des données pertinentes pour l'organisation.	Une personne qui utilise les données pour obtenir des résultats opérationnels
Résumé des responsabilités	<ul style="list-style-type: none"> ▪ Conformité ▪ Risque ▪ Surveillance ▪ Approbation ▪ Champion ▪ Résolution des problèmes 	<ul style="list-style-type: none"> ▪ Précision ▪ Cohérence ▪ Exigences opérationnelles ▪ Définition et gestion des métadonnées ▪ Qualité des données ▪ Curation des données ▪ Aptitude à l'emploi ▪ Gouvernance ▪ Inventaire ▪ RDM ▪ Gestion des rôles 	<ul style="list-style-type: none"> ▪ Sécurité ▪ Gestion de l'accès ▪ Disponibilité ▪ Capacité ▪ Continuité ▪ Sauvegarde ▪ Mise en œuvre ▪ Normes techniques ▪ Configuration ▪ Contrôle ▪ Modélisation ▪ Versionnage ▪ Gestion des changements 	<ul style="list-style-type: none"> ▪ Acquisition et saisie des données ▪ Qualité des données ▪ Préparation des métadonnées ▪ Collecte éthique et sécurisée des données ▪ Identification des problèmes ▪ Identification de nouvelles sources de données 	<ul style="list-style-type: none"> ▪ Utilisation éthique ▪ Rapport sur la qualité des données ▪ Rapport sur l'aptitude à l'emploi ▪ Identification des règles de gestion et de données ▪ Identification et rapport sur le contrôle des données ▪ Utilisation conforme à la gouvernance

Objectifs de la gouvernance des données

1

Créer une culture de données libre-service

2

Établir des règles internes pour l'utilisation des données

3

Mettre en œuvre les exigences de conformité

4

Améliorer les communications internes et externes

5

Accroître la valeur des données

6

Réduire les coûts

7

Gérer continuellement les risques

8

Assurer la pérennité de l'entreprise

Problèmes communs à la gouvernance

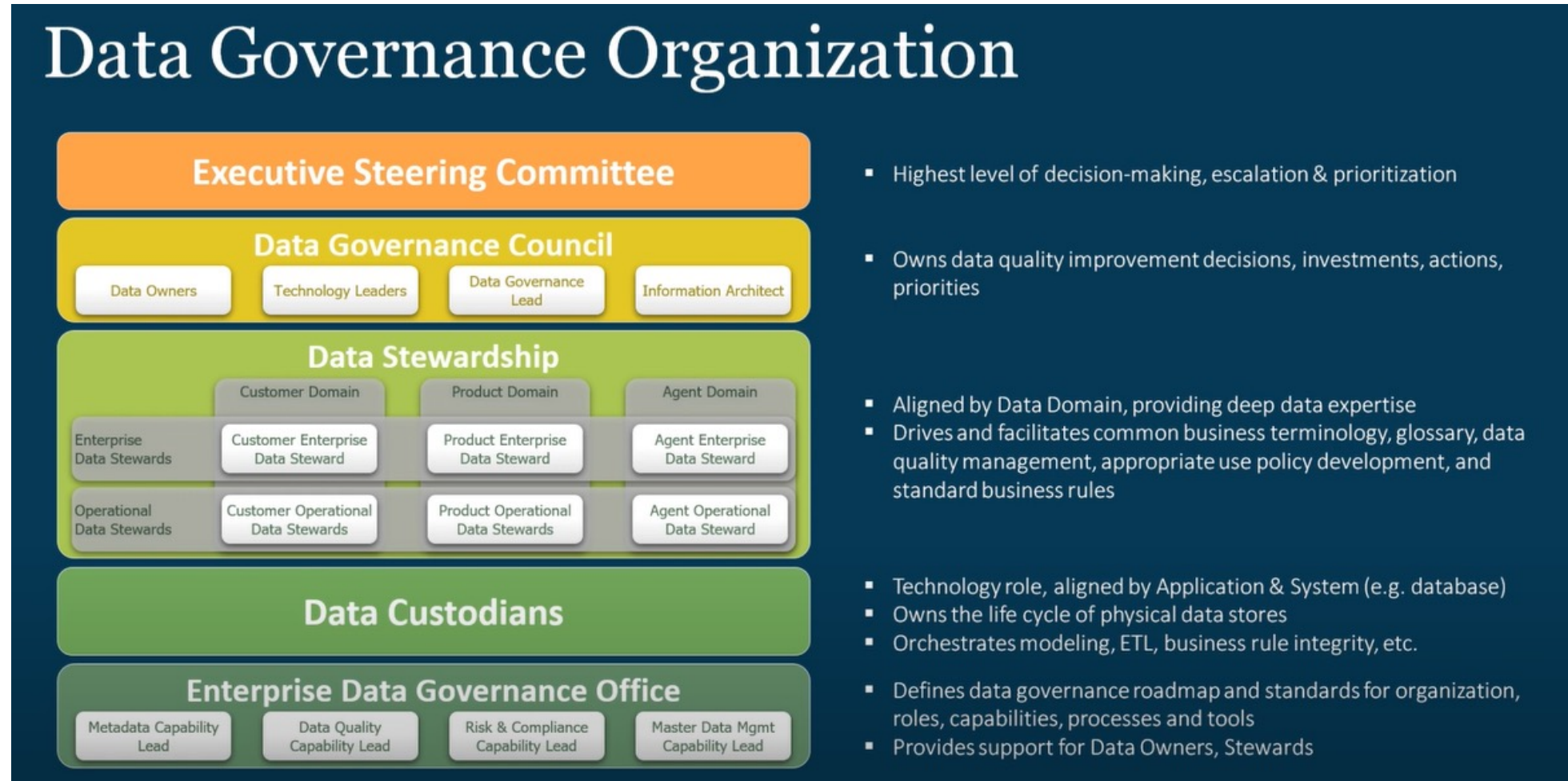
Pionnier, planifié et projeté par les ressources les plus techniques

Le verbiage est partout.

La perception des "réunions de beignets".



Structure possible



Évaluation ad hoc : Reconnaissance des barrières

Acquisition de données



Acquisition de données

Sources de données multiples

Processus manuel

Aucune norme

Pas de processus défini

Pas de validation des données

Maintenance des données



Principalement manuel

Problèmes d'intégration des données

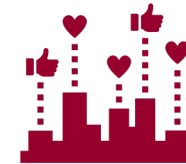
Manque de nettoyage des données

Duplication des données

Peu ou pas d'accessibilité

Aucun dictionnaire de données

Diffusion des données



La bureaucratie informatique constante

Le nettoyage des données mal ordonné

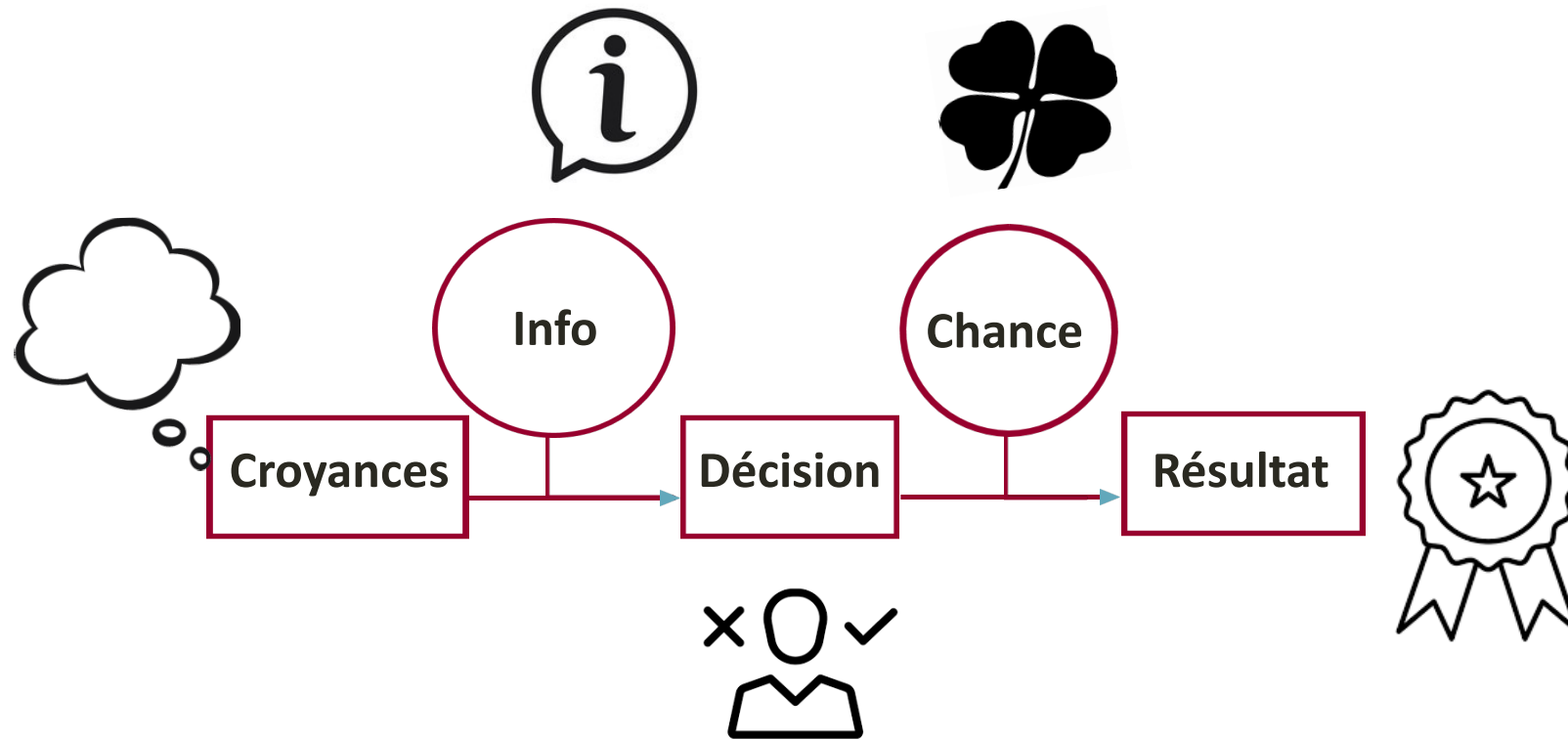
Données non cohérentes

Changements non communiqués

La culture de l'épaulement

Rapports envoyés par courriel/chat

Chance et information (reprise)



Références suggérées

Storytelling with Data, by Cole Nussbaumer Knaflic

Weapons of Math Destruction, by Cathy O'Neil

How to Decide: Simple Tools for Making Better Choices, by Annie Duke

The Signal and the Noise, by Nate Silver

Superforecasting: The Art and Science of Prediction, by Philip E. Tetlock

Business Analytics: Data Analysis & Decision-Making, by Albright and Winston

Data Understanding, Data Analysis, and Data Science, by P. Boily and J. Schellinck

Table ronde



Déverrouiller la décision