

Data Science : Communication

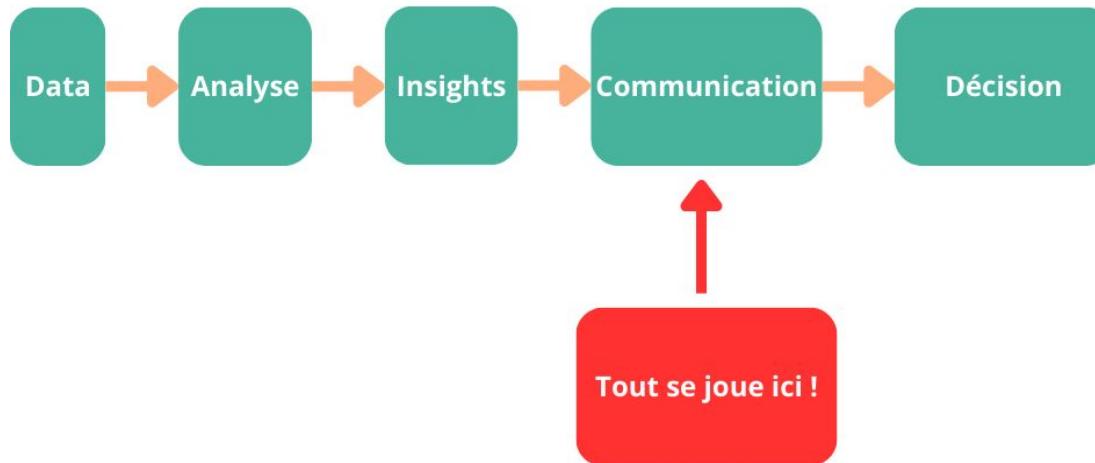
J. Delpech – Cours : EDA
Cursus B3 Data/IA – 2025/2026
Dernière mise à jour : février 2026

1. Introduction : pourquoi la communication est cruciale

1.1 Le paradoxe du Data Scientist

"Le meilleur modèle du monde ne sert à rien si personne ne comprend ce qu'il dit."

Le Data Scientist passe 80% de son temps sur l'analyse... mais **100% de l'impact** dépend de sa capacité à communiquer les résultats.



1.2 Les enjeux d'une bonne communication

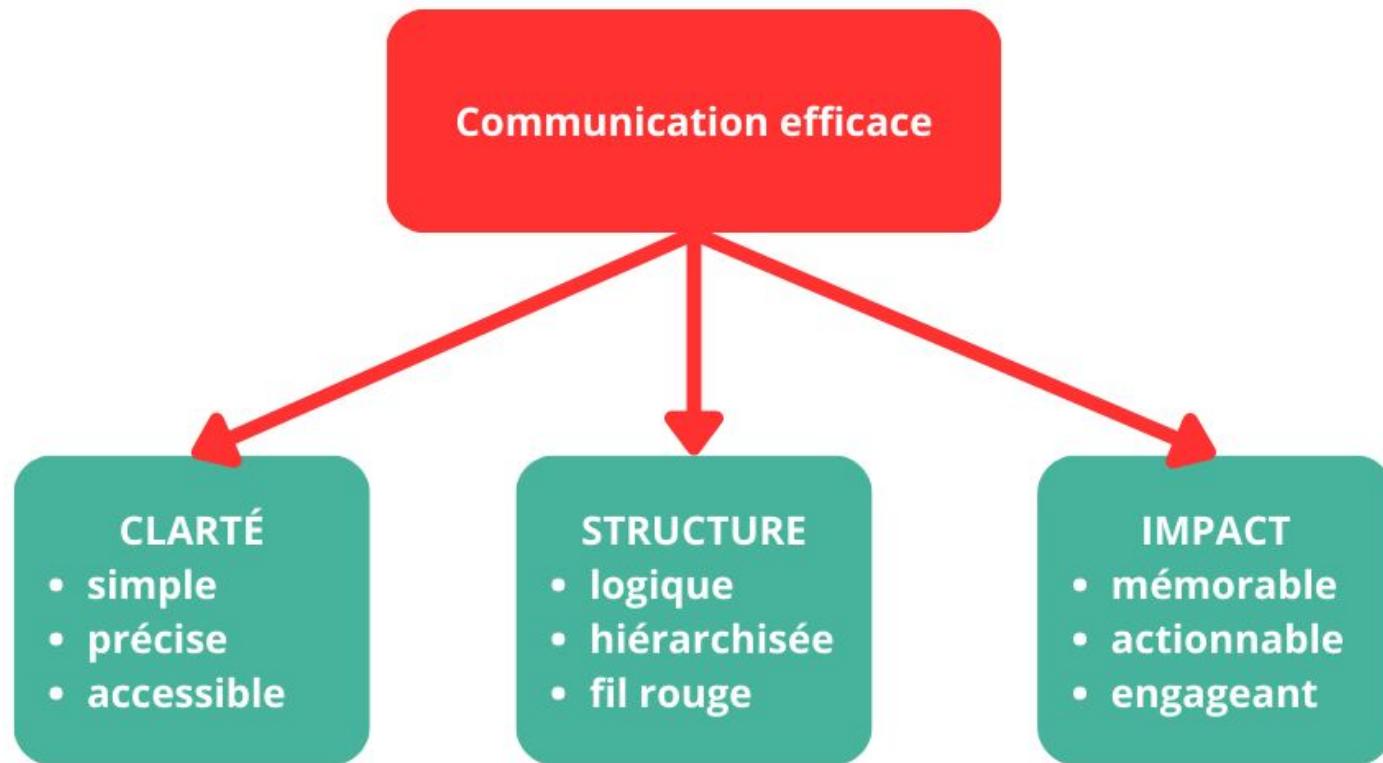
| Bonne communication | Mauvaise communication |
|--------------------------------|---|
| Décisions éclairées et rapides | Paralysie décisionnelle |
| Confiance dans la data | Méfiance envers les analyses |
| Adoption des recommandations | Rejet ou ignorance des insights |
| Crédibilité du Data Scientist | Perception de "tour d'ivoire technique" |
| Impact business mesurable | Projets qui finissent dans un tiroir |

1.3 Les défis spécifiques à la Data Science

La communication en Data Science présente des défis uniques :

- **Complexité technique** : expliquer des algorithmes sans perdre l'audience
- **Incertitude** : communiquer des probabilités, des intervalles de confiance
- **Volume** : synthétiser des analyses massives en messages clairs
- **Audiences variées** : du comité de direction à l'équipe technique
- **Biais cognitifs** : le public peut mal interpréter les statistiques

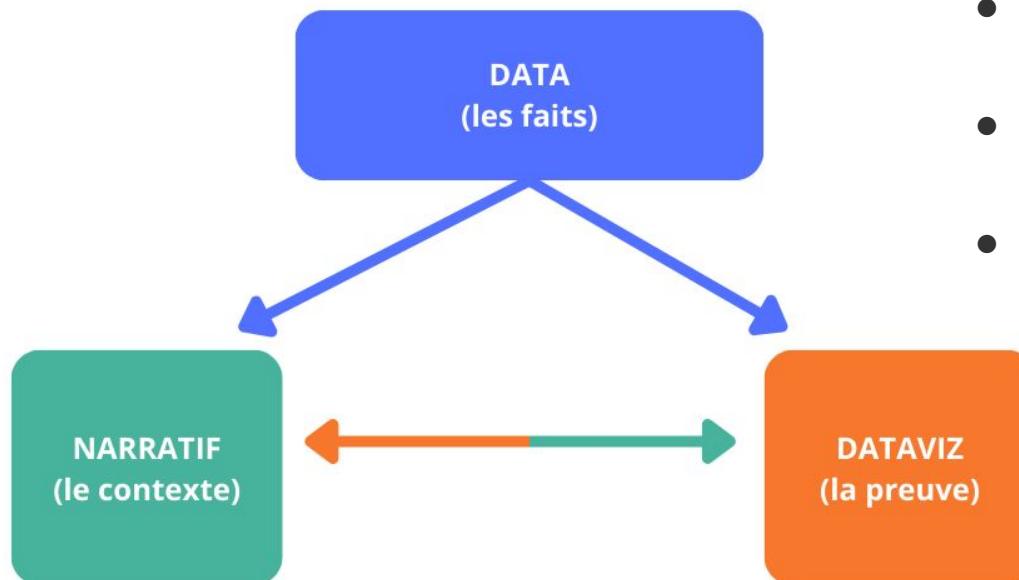
1.4 Les trois piliers de la communication data



2. Data Storytelling : raconter une histoire avec les données

2.1 Qu'est-ce que le Data Storytelling ?

Le Data Storytelling est l'art de **transformer des données en récit** pour informer, convaincre et inspirer l'action.



- **Data** : les chiffres, les analyses, les insights
- **Narrative** : le contexte, l'histoire, le "pourquoi ça compte"
- **Visualisation** : les graphiques, les représentations visuelles

Seules, les données informent.

Les données + une histoire transforment.

2.2 Pourquoi le storytelling fonctionne

Notre cerveau est câblé pour les histoires :

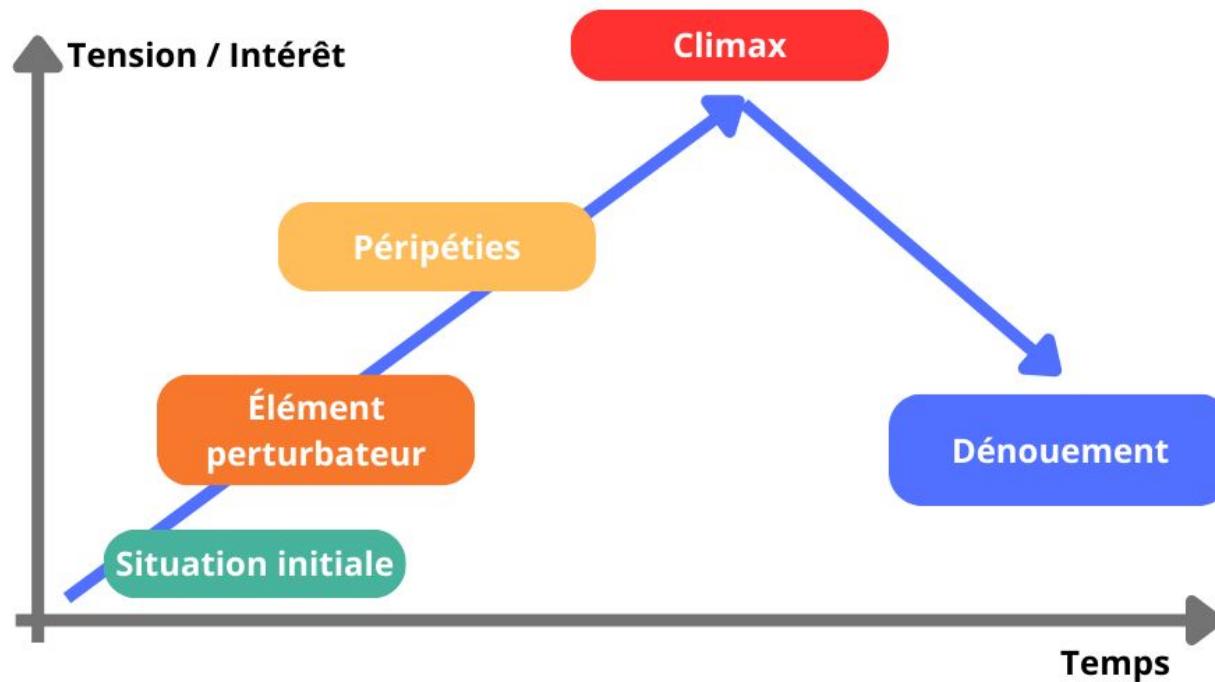
| Information brute | Information narrativisée |
|-----------------------------------|--|
| Cognition « froide » (analytique) | Active le traitement émotionnel (cognition « chaude ») |
| Oubliée en quelques minutes | Retenue 22x plus longtemps (études Stanford) |
| Génère de l'indifférence | Crée de l'empathie et de l'engagement |
| "Intéressant..." | "Il faut agir !" |

Exemple :

- Ne pas se contenter de dire : "Le taux de churn est de 15% ce trimestre"
- Dire plutôt : "Chaque mois, nous perdons l'équivalent de la ville de Bordeaux en clients. Et 60% d'entre eux partent pour la même raison : un temps de réponse trop long du service client."

2.3 La structure narrative classique

Toute bonne histoire suit un arc narratif. Adaptons-le à la data :



2.3 La structure narrative classique

| Acte | Contenu | Exemple |
|--|------------------------------|--|
| 1. Contexte = Situation initiale | Situation initiale, enjeux | "Notre objectif 2024 est de réduire le churn de 20%" |
| 2. Problème = élément perturbateur | Ce qui ne va pas, la tension | "Or, le churn a augmenté de 3 points ce trimestre" |
| 3. Exploration / analyse = péripéties | L'analyse, les découvertes | "En analysant les données, nous avons identifié..." |
| 4. Insight = climax | La révélation principale | "80% des churners ont contacté le support 3+ fois" |
| 5. Action = dénouement | Les recommandations | "En améliorant le support, nous pouvons réduire le churn de 12%" |

2.4 Les ingrédients d'un bon Data Story

1. Un héros (votre audience ou l'entreprise)

"Votre équipe commerciale passe 4h par jour sur des tâches administratives..."

2. Un enjeu clair (pourquoi ça compte)

"...ce qui représente 500K€ de coût d'opportunité annuel."

3. Un conflit/une tension (le problème à résoudre)

"Le système actuel ne permet pas d'identifier les opportunités prioritaires."

4. Une découverte (l'insight data)

"Notre analyse montre que 3 signaux prédisent 80% des conversions."

5. Une résolution (l'action recommandée)

"Un outil de scoring permettrait de concentrer les efforts sur les bons prospects."

2.5 Les 3 types de Data Stories

Comme il y a des romans initiatiques, d'aventures, etc. il y a différentes data stories

| Type | Objectif | Structure | Exemple |
|--------------|---------------------------|------------------------------|---|
| Explicatif | Informer, éduquer | Linéaire, pédagogique | "Voici comment a évolué notre marché en 2024" |
| Exploratoire | Susciter des questions | Interactif, ouvert | "Que nous disent les données sur nos clients ?" |
| Persuasif | Convaincre, faire agir | Argumentatif, orienté action | "Pourquoi nous devons investir dans le segment B2B" |

2.6 Erreurs courantes en Data Storytelling

| Erreur | Problème | Solution |
|---------------------------------|----------------------------|-----------------------------------|
| Tout montrer | Noie le message principal | Sélectionner impitoyablement |
| Commencer par la méthodo | Perd l'audience d'entrée | Commencer par l'enjeu business |
| Pas de "so what?" | Insights sans implications | Toujours conclure par l'action |
| Jargon technique | Exclut les non-experts | Traduire en langage métier |
| Données sans contexte | Impossible d'interpréter | Comparer, benchmarker, historiser |

3. La méthode STAR : structurer ses messages

3.1 Présentation de la méthode

La méthode **STAR** est un framework de communication structurée, originellement utilisé pour les entretiens d'embauche, mais très efficace pour présenter des résultats d'analyse.

| Étape | Contenu |
|---------------|--|
| S / Situation | Le contexte, le point de départ |
| T / Task | L'objectif, ce qu'il fallait accomplir |
| A / Action | Ce que vous avez fait |
| R / Result | Les résultats obtenus, l'impact |

3.2 Application en Data Science

Exemple : présenter un projet de détection de fraude

| Composante | Contenu |
|------------|---|
| Situation | "L'entreprise perdait 2M€ par an en fraudes non détectées. Le système de règles existant ne détectait que 30% des cas." |
| Task | "L'objectif était d'améliorer le taux de détection tout en limitant les faux positifs qui génèrent des frictions client." |
| Action | "J'ai développé un modèle de machine learning en analysant 2 ans d'historique de transactions. J'ai testé 3 approches et retenu un Random Forest optimisé." |
| Result | "Le nouveau modèle détecte 75% des fraudes (+45 points) avec seulement 2% de faux positifs. Économie estimée : 1,2M€/an." |

3.3 Variante STAR-L (avec Learnings)

Pour les présentations de retour d'expérience, ajoutez les **apprentissages** :

S → T → A → R → L (Learnings)

L - Learnings : "Ce projet m'a appris que l'implication du métier dès le départ est cruciale. Les règles de fraude qu'ils connaissaient ont été intégrées comme features et ont significativement amélioré le modèle."

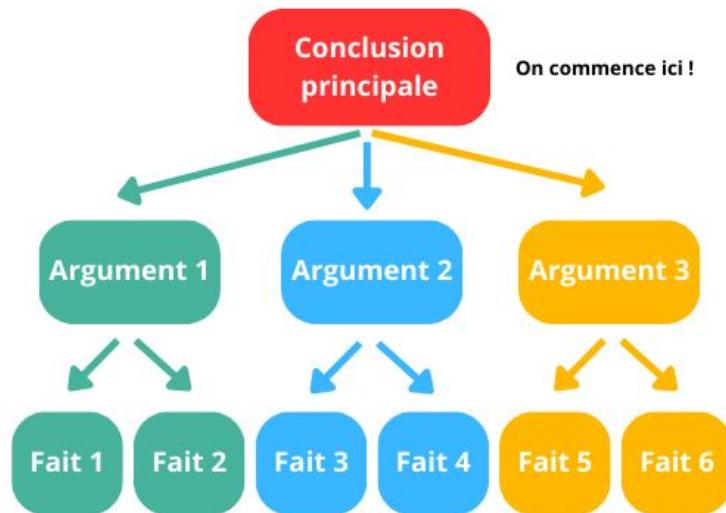
3.4 Quand utiliser STAR

- Présentation de résultats de projet
- Entretiens d'embauche (questions comportementales)
- Rapports d'analyse à la direction
- Documentation de cas d'usage
- Retours d'expérience (RETEX)

4. La Pyramide de Minto : structurer pour convaincre

4.1 Le principe

La **Pyramide de Minto** (ou Pyramid Principle), développée par Barbara Minto chez McKinsey, propose de structurer la communication en **commençant par la conclusion**.



4.2 Avantage

Approche classique (bottom-up) :

"J'ai analysé les données... puis j'ai fait tel traitement... puis j'ai découvert que... et donc ma conclusion est..."

Problème : l'audience doit attendre la fin pour comprendre où vous allez. Elle décroche.

Approche Minto (top-down) :

"Ma recommandation est X. Voici les 3 raisons qui la soutiennent. Détaillons chacune..."

Avantage : l'audience sait immédiatement l'essentiel et peut choisir de creuser ou non.

4.3 Les règles de la Pyramide

Règle 1 : Commencer par la réponse

- Votre conclusion/recommandation en premier
- Le "so what?" tout de suite

Règle 2 : Grouper par 3 (± 2)

- Maximum 5 arguments par niveau (idéalement 3)
- Au-delà, l'audience ne retient plus

Règle 3 : Logique MECE

- Mutually Exclusive : pas de chevauchement entre arguments
- Collectively Exhaustive : couverture complète du sujet

Règle 4 : Cohérence verticale et horizontale

- Vertical : chaque niveau supporte celui du dessus
- Horizontal : les éléments d'un même niveau sont de même nature

4.4 Application en Data Science

Exemple : recommandation d'investissement marketing

CONCLUSION : Nous devons réallouer 30% du budget TV vers le digital.

- **ARGUMENT 1 : *Le digital a un ROI 3x supérieur***
 - Fait : ROI digital = 4,2 vs ROI TV = 1,4
 - Fait : Coût d'acquisition digital = 12€ vs TV = 45€
- **ARGUMENT 2 : *Notre cible est majoritairement digitale***
 - Fait : 78% des 25-45 ans passent +3h/jour en ligne
 - Fait : Audience TV en baisse de 15% sur notre cible
- **ARGUMENT 3 : *Le digital permet une optimisation continue***
 - Fait : A/B testing possible en temps réel
 - Fait : Attribution mesurable à la conversion

4.5 Le "Elevator Pitch" avec Minto

La pyramide permet de s'adapter au temps disponible :

| Temps | Niveau de détail |
|--------------------|-------------------------------------|
| 30 secondes | Conclusion seule |
| 2 minutes | Conclusion + 3 arguments |
| 10 minutes | Conclusion + arguments + faits clés |
| 30 minutes | Structure complète avec détails |

5. La méthode SCR :

Situation - Complication - Résolution

5.1 Présentation

La méthode **SCR** est une structure narrative simple et percutante, particulièrement adaptée aux présentations exécutives.

| Étape méthode SCR | Contenu |
|-------------------|--|
| S / Situation | L'état des lieux, le contexte stable |
| C / Complication | Le problème, ce qui change, la tension |
| R / Résolution | La solution, la recommandation |

5.2 Efficacité

La structure SCR crée naturellement une **tension** qui capte l'attention :

1. **Situation** : pose le décor, crée un terrain commun
2. **Complication** : introduit un déséquilibre, une urgence
3. **Résolution** : apporte le soulagement, la solution

C'est le schéma de base de toute histoire captivante.

5.3 Application en Data Science

Exemple : présentation d'une analyse de churn

| | |
|---------------------|---|
| S - Situation | "Notre base clients a crû de 40% en 2024. Nous avons atteint 500K clients actifs, un record historique." |
| C - Complication | "Cependant, notre taux de churn a atteint 18%, son plus haut niveau. Nous perdons des clients plus vite que nous n'en gagnons. À ce rythme, notre base nette sera négative d'ici Q3." |
| R - Résolution | "Notre analyse révèle que 3 actions ciblées peuvent réduire le churn de 40% : améliorer l'onboarding, créer un programme de fidélité, et réduire les temps de réponse support." |

5.4 Variante SCQA (avec Question)

Pour les présentations plus longues, insérez une **Question** entre la Complication et la Résolution :

S → C → Q → A (Answer/Résolution)

Q - Question : "Comment pouvons-nous inverser cette tendance et fidéliser nos clients ?"

Cela permet de marquer une pause et de recentrer l'attention avant la solution.

6. La méthode du Qui ? Quand ? Quoi ?... : structurer l'information

6.1 Présentation

Le **Q**Q**OQCCP ou les **5W + H**** constituent une méthode journalistique pour s'assurer de couvrir tous les aspects d'un sujet.

| 5W + H | Couverture |
|---------------|---|
| WHO? | Qui est concerné ? Qui agit ? |
| WHAT? | De quoi s'agit-il ? Quel est le problème/solution ? |
| WHERE? | Où cela se passe-t-il ? Quel périmètre ? |
| WHEN? | Quand ? Quelle temporalité ? Quels délais ? |
| WHY? | Pourquoi ? Quelles causes ? Quels enjeux ? |
| HOW? | Comment ? Quelle méthode ? Quels moyens ? |

6.2 Application en Data Science

Check-list pour une présentation de résultats :

| Question | À couvrir |
|----------|--|
| WHO | Qui est impacté ? Qui doit agir ? Qui a commandité l'analyse ? |
| WHAT | Quel insight principal ? Quelle recommandation ? |
| WHERE | Quel segment ? Quelle région ? Quel produit ? |
| WHEN | Sur quelle période ? Quelle urgence ? Quel planning ? |
| WHY | Pourquoi ce résultat ? Pourquoi agir maintenant ? |
| HOW | Comment avez-vous analysé ? Comment mettre en œuvre ? |

6.3 Exemple d'application

Présentation d'une anomalie détectée :

WHAT : Une baisse anormale de 25% des ventes a été détectée.

WHERE : Le phénomène touche exclusivement la région Sud-Est.

WHEN : La baisse a commencé il y a 3 semaines, le 15 janvier.

WHO : Principalement les clients B2B du segment PME.

WHY : L'analyse montre une corrélation avec le départ de 2 commerciaux clés et l'arrivée d'un concurrent agressif.

HOW : Je recommande un plan d'action en 3 points : recrutement urgent, contre-offre commerciale, suivi renforcé des clients à risque.

7. Adapter sa communication à l'audience

7.1 Les profils d'audience

| Audience | Ce qui l'intéresse | Ce qui l'ennuie | Format adapté |
|-------------------------------|--|----------------------------------|------------------------------------|
| Direction générale | Impact business, ROI, risques, décisions | Détails techniques, méthodologie | Executive summary, 5 slides max |
| Managers opérationnels | Actions concrètes, planning, ressources | Théorie, complexité excessive | Recommandations actionnables |
| Équipes métier | Utilité au quotidien, facilité d'usage | Jargon data, abstractions | Démos, exemples concrets |
| Équipes techniques | Méthodologie, robustesse, code | Vulgarisation excessive | Documentation technique |
| Clients externes | Valeur, confiance, simplicité | Incertitudes, complexité | Messages clairs, preuves visuelles |

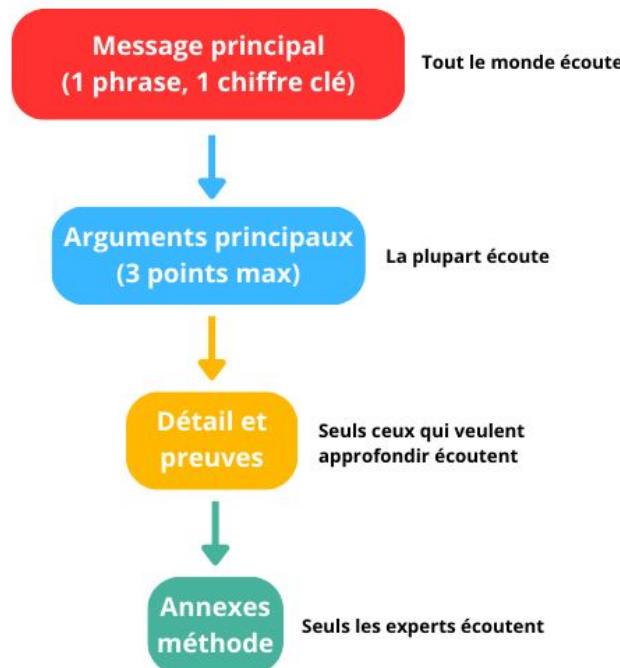
7.2 La règle du "So What?"

Pour chaque information présentée, posez-vous la question : "**Et alors ?**"

| Information brute | So What? (pour un DG) | So What? (pour un manager) |
|-----------------------------------|---|--|
| "Le modèle a une accuracy de 87%" | "Cela permet de réduire les coûts de 15%" | "L'équipe peut traiter 30% de cas en plus" |
| "J'ai utilisé un Random Forest" | (Ne pas mentionner) | "C'est un modèle robuste et explicable" |
| "Le R ² est de 0.82" | "Les prévisions sont fiables à 80%" | "On peut se fier aux recommandations" |

7.3 Adapter le niveau de détail

Principe de l'entonnoir inversé : commencez large, proposez d'approfondir.



7.4 Les questions à anticiper par audience

| Audience | Questions probables |
|-----------|--|
| DG/Comex | Quel ROI ? Quels risques ? Quel investissement ? Quel délai ? |
| Finance | Comment avez-vous calculé ? Quelles hypothèses ? Quelle marge d'erreur ? |
| Juridique | C'est conforme RGPD ? Quels risques légaux ? |
| IT | Comment ça s'intègre ? Quelles ressources ? Qui maintient ? |
| Métier | Ça va changer quoi pour moi ? C'est fiable ? Comment je l'utilise ? |

8. Visualisation des données : les fondamentaux

8.1 Pourquoi visualiser ?

La DataViz est une composante essentielle du storytelling (= démonstration)

"Une image vaut mille mots. Un bon graphique vaut mille lignes de données."

La visualisation permet de :

- **Révéler** des patterns invisibles dans les chiffres
- **Comparer** rapidement des ordres de grandeur
- **Mémoriser** l'information (réception visuelle)
- **Convaincre** avec une preuve immédiate

8.2 Choisir le bon graphique

| Objectif | Type de graphique | Exemple d'usage |
|--------------------------|-------------------------------------|------------------------|
| Comparer des catégories | Barres (horizontales ou verticales) | Ventes par région |
| Montrer une évolution | Ligne, aire | CA mensuel sur 3 ans |
| Montrer une distribution | Histogramme, box plot | Répartition des âges |
| Montrer une proportion | Camembert (max 5 parts), treemap | Parts de marché |
| Montrer une corrélation | Nuage de points (scatter plot) | Prix vs surface |
| Montrer une composition | Barres empilées, waterfall | Décomposition du coût |
| Montrer une géographie | Carte choroplète | Ventes par département |

8.3 Les règles d'or de la dataviz

1. Un graphique = un message

- Pas de graphique "fourre-tout"
- Titre (slide) = le message, pas la description

| Mauvais titre | Bon titre |
|----------------------------------|---------------------------------------|
| "Évolution des ventes 2020-2024" | "Les ventes ont doublé en 4 ans" |
| "Répartition par segment" | "Le segment B2B représente 60% du CA" |

8.3 Les règles d'or de la dataviz

2. Réduire le "chart junk"

- Supprimer les éléments décoratifs inutiles
- Pas de 3D (déforme la perception)
- Pas de couleurs arc-en-ciel (penser accessibilité)
- Grilles légères ou absentes

8.3 Les règles d'or de la dataviz

3. Guider le regard

- Mettre en évidence le point clé (couleur, annotation)
- Ordonner les données de façon logique (croissant/décroissant)
- Utiliser les annotations pour expliquer

8.3 Les règles d'or de la dataviz

4. Respecter les conventions

- Axe Y commence à 0 pour les barres
- Le temps va de gauche à droite
- Rouge = négatif/danger, Vert = positif/succès

8.4 Exemple de transformation

Avant (surcharge d'information) :

Tableau de 50 lignes avec 12 colonnes de KPIs mensuels...

→ Impossible à lire, aucun message ne ressort

Après (message clair) :

Un line chart avec 2 courbes (réel vs objectif)

+ Une annotation sur le point d'inflexion

+ Titre : "Le retard sur objectif se creuse depuis mars"

→ Message immédiat, action évidente

9. Gérer l'incertitude dans la communication

9.1 Le défi de l'incertitude

La Data Science produit des **probabilités**, pas des certitudes. Or, les décideurs veulent des réponses claires.

| Ce que dit le Data Scientist | Ce qu'entend le décideur |
|--|---|
| "Le modèle prédit avec 75% de probabilité" | "Donc ça va arriver" (ou "Donc on ne sait pas") |
| "L'intervalle de confiance est [10, 50]" | "C'est 30 alors ?" |
| "Sous réserve des hypothèses..." | "Il n'est pas sûr de lui" |

9.2 Comment communiquer l'incertitude

1. Contextualiser avec des comparaisons

"Notre modèle est correct 8 fois sur 10. C'est mieux que les experts humains (6/10) et suffisant pour filtrer 70% des cas simples."

2. Utiliser des scénarios

"Dans le scénario optimiste, le gain sera de 500K€. Dans le scénario pessimiste, 200K€. Le plus probable est autour de 350K€."

3. Visualiser l'incertitude

- Barres d'erreur
- Intervalles de confiance ombrés
- Distributions (fan charts)

4. Être transparent sur les limites

"Cette analyse suppose que les tendances passées se poursuivent. Un événement majeur (nouveau concurrent, crise) changerait la donne."

9.3 Les formulations à privilégier

| Éviter les formulations floues | Privilégier une formulation claire |
|------------------------------------|--|
| "Le modèle est assez précis" | "Le modèle se trompe dans 15% des cas" |
| "Il y a un risque" | "Le risque est de 20%, soit 1 cas sur 5" |
| "C'est probable" | "Il y a 3 chances sur 4 que cela se produise" |
| "Les résultats sont significatifs" | "L'effet observé a moins de 5% de chances d'être dû au hasard" |

10. Structurer une présentation data

10.1 Template de présentation (10-15 slides)

| Slide | Étape | Contenu / Exemple |
|------------|---|---|
| SLIDE 1 | Titre + accroche | "Réduire le churn de 30% : c'est possible" |
| SLIDE 2 | Executive Summary (la conclusion d'abord) | Message clé + 3 points + recommandation |
| SLIDE 3 | Contexte et enjeux (Situation) | Pourquoi ce sujet ? Quel impact business ? |
| SLIDE 4 | Le problème (Complication) | Qu'est-ce qui ne va pas ? Données clés |
| SLIDES 5-8 | L'analyse (3-4 slides) | Insights principaux, un par slide / Graphiques clairs avec messages |
| SLIDE 9 | Synthèse des découvertes | Récapitulatif visuel des insights |
| SLIDE 10 | Recommandations (Résolution) | Actions concrètes, priorisées |
| SLIDE 11 | Impact attendu et ROI | Chiffres, scénarios, délais |
| SLIDE 12 | Prochaines étapes | Qui fait quoi, quand, prochaine réunion |
| ANNEXES | | Méthodologie, détails techniques, données brutes (pour les questions) |

10.2 La règle du 10-20-30

Popularisée par Guy Kawasaki :

- **10 slides** maximum (hors annexes)
- **20 minutes** de présentation
- **30 points** minimum pour la taille de police

10.3 Check-list avant présentation

Contenu :

- Le message principal est-il clair en 1 phrase ?
- Chaque slide a-t-il un seul message ?
- Les "so what?" sont-ils explicites ?
- Les recommandations sont-elles actionnables ?
- Les annexes sont-elles prêtes pour les questions ?

Forme :

- Les graphiques sont-ils lisibles (taille, contraste) ?
- Le jargon technique est-il éliminé ou expliqué ?
- La présentation tient-elle dans le temps imparti ?
- Les transitions sont-elles fluides ?

Audience :

- Ai-je adapté le niveau de détail à mon audience ?
- Ai-je anticipé les questions probables ?
- Les décideurs peuvent-ils agir avec ces informations ?