

Manuel de Préparation: Visualisation de Données

Introduction

La visualisation de données est l'art de représenter graphiquement des informations pour faciliter leur compréhension, leur analyse et la prise de décision. Dans le contexte bancaire, une bonne visualisation permet de communiquer efficacement avec les stakeholders et d'identifier rapidement des tendances ou anomalies.

1. Principes Fondamentaux

1.1 Objectifs de la Visualisation

- **Explorer:** Découvrir des patterns dans les données
- **Expliquer:** Communiquer des insights à un public
- **Confirmer:** Valider des hypothèses
- **Alerter:** Signaler des anomalies ou risques

1.2 Règles d'Or

1. **Clarté:** Un graphique = Un message
 2. **Honnêteté:** Ne pas déformer les données
 3. **Pertinence:** Choisir le bon type de graphique
 4. **Simplicité:** Éviter le "chartjunk" (éléments inutiles)
 5. **Contexte:** Toujours fournir les unités et références
-

2. Types de Variables et Graphiques Appropriés

2.1 Taxonomie des Variables

Type	Description	Exemples Bancaires
Nominale	Catégories sans ordre	Type de compte, agence, produit
Ordinale	Catégories ordonnées	Score de risque (A, B, C), satisfaction client
Discrète	Valeurs entières comptables	Nombre de transactions, nombre de comptes
Continue	Valeurs mesurables	Solde, montant de prêt, taux d'intérêt

2.2 Matrice de Sélection des Graphiques

Objectif	1 Variable	2 Variables	3+ Variables
Distribution	Histogramme, Box plot	Scatter plot	Heatmap
Comparaison	Bar chart	Grouped bar	Stacked bar
Composition	Pie chart	Stacked bar	Treemap
Relation	-	Scatter plot	Bubble chart
Tendance	Line chart	Multi-line	Area chart

3. Graphiques en Détail

3.1 Histogramme

Définition Représentation de la distribution d'une variable continue en regroupant les valeurs en intervalles (bins).

Statistiques Associées

- **Forme:** Symétrique, asymétrique (skewness)
- **Centre:** Où se concentrent les données (mode)
- **Dispersion:** Étendue des données
- **Outliers:** Valeurs extrêmes isolées

Quand l'utiliser

- Comprendre la distribution d'une variable continue
- Détecter la normalité des données
- Identifier les valeurs aberrantes

Contexte Bancaire

Exemple: Distribution des montants de prêts

- Permet de voir si la banque accorde majoritairement des petits ou gros prêts
- Identifie les segments de marché
- Détecte des concentrations de risque

Interprétation Statistique

Distribution normale: Symétrique, en cloche

- Moyenne Médiane Mode

Distribution asymétrique droite (positive skew):

- Mode < Médiane < Moyenne
- Typique des revenus, montants de transactions

Distribution asymétrique gauche (negative skew):

- Moyenne < Médiane < Mode
- Plus rare, ex: scores de satisfaction élevés

Erreurs à Éviter

- Nombre de bins mal choisi (trop peu = perte d'info, trop = bruit)
 - Règle de Sturges: $k = 1 + 3.322 \times \log_{10}(n)$
 - Ou utiliser la règle de Freedman-Diaconis
-

3.2 Box Plot (Boîte à Moustaches)

Définition Représentation des 5 statistiques sommaires: min, Q1, médiane, Q3, max.

Composants

- ← Q3 (75ème percentile)
- ← Médiane (Q2, 50ème percentile)
- ← Q1 (25ème percentile)
- ← Moustaches ($1.5 \times \text{IQR}$)
- ← Outliers (au-delà des moustaches)

Statistiques Associées

- **IQR (Interquartile Range):** $Q3 - Q1$
- **Outliers:** $< Q1 - 1.5 \times \text{IQR}$ ou $> Q3 + 1.5 \times \text{IQR}$
- **Skewness:** Position de la médiane dans la boîte

Quand l'utiliser

- Comparer des distributions entre groupes
- Identifier rapidement les outliers
- Visualiser la dispersion et la symétrie

Contexte Bancaire

Exemple: Comparaison des soldes par type de compte

- Compte épargne vs compte courant vs compte entreprise
- Permet de voir les différences de distribution
- Identifie les comptes avec des soldes anormaux

Interprétation

Médiane proche de Q3: Distribution asymétrique gauche
Médiane proche de Q1: Distribution asymétrique droite
Boîte petite: Données concentrées
Moustaches longues: Grande variabilité
Beaucoup d'outliers: Investiguer la qualité des données

3.3 Bar Chart (Diagramme en Barres)

Définition Représentation de quantités par des barres rectangulaires proportionnelles.

Types

- **Simple:** Une catégorie, une mesure
- **Groupé:** Comparaison de sous-groupes
- **Empilé:** Composition d'un total

Quand l'utiliser

- Comparer des valeurs entre catégories
- Visualiser des fréquences ou des comptages
- Montrer des rankings

Contexte Bancaire

Exemple: Nombre de nouveaux comptes par agence

- Classement des performances des agences
- Identification des agences sous-performantes
- Base pour allocation des ressources

Bonnes Pratiques

- Toujours commencer l'axe Y à zéro
 - Ordonner les barres de façon logique (alphabétique, valeur)
 - Limiter à 5-7 catégories maximum
 - Utiliser l'orientation horizontale si labels longs
-

3.4 Line Chart (Graphique Linéaire)

Définition Représentation de l'évolution d'une variable continue sur un axe temporel ou ordonné.

Statistiques Associées

- **Tendance:** Croissante, décroissante, stable
- **Saisonnalité:** Patterns récurrents
- **Cyclicité:** Variations à long terme
- **Volatilité:** Amplitude des variations

Quand l'utiliser

- Montrer l'évolution dans le temps
- Identifier des tendances
- Comparer des séries temporelles

Contexte Bancaire

Exemple: Évolution du NPL ratio mensuel

- Identifie les tendances de détérioration
- Permet de corrélérer avec des événements
- Base pour les projections

Interprétation Statistique

Trend analysis:

- Moyenne mobile pour lisser les variations
- Régression linéaire pour quantifier la tendance

Saisonnalité bancaire:

- Fin de mois: pic de transactions

- Fin d'année: clôtures comptables
 - Périodes de fêtes: hausse des retraits
-

3.5 Scatter Plot (Nuage de Points)

Définition Représentation de la relation entre deux variables continues.

Statistiques Associées

- **Corrélation (r):** Force et direction de la relation
 - $r = 1$: Corrélation positive parfaite
 - $r = 0$: Pas de corrélation linéaire
 - $r = -1$: Corrélation négative parfaite
- **R^2 :** Proportion de variance expliquée

Quand l'utiliser

- Explorer la relation entre deux variables
- Identifier des clusters
- Détecter des outliers bivariés

Contexte Bancaire

Exemple: Revenu vs Montant de prêt accordé

- Visualise la politique de crédit
- Identifie des cas hors norme
- Permet d'évaluer le risque de concentration

Interprétation

Pattern linéaire positif: Quand X augmente, Y augmente

Pattern linéaire négatif: Quand X augmente, Y diminue

Pattern non-linéaire: Relation curvilinéaire (log, polynomial)

Pas de pattern: Variables indépendantes

Clusters: Groupes distincts dans les données

Attention: Corrélation \neq Causalité

Exemple bancaire:

- Corrélation entre nombre de cartes de crédit et défaut
 - Ne signifie pas que les cartes causent le défaut
 - Variable confondante possible: comportement financier
-

3.6 Pie Chart (Diagramme Circulaire)

Définition Représentation de la composition d'un tout en parts proportionnelles.

Quand l'utiliser (avec prudence)

- Montrer des proportions d'un total
- Maximum 5-6 catégories
- Quand les parts sont significativement différentes

Quand NE PAS l'utiliser

- Comparaison précise de valeurs (utiliser bar chart)
- Beaucoup de catégories
- Évolution dans le temps
- Valeurs négatives

Contexte Bancaire

Exemple: Répartition du portefeuille de prêts par secteur

- Agriculture: 15%
- Commerce: 40%
- Industrie: 25%
- Services: 20%

Alternative Recommandée Le **donut chart** ou mieux, le **bar chart horizontal** est souvent plus lisible.

3.7 Heatmap (Carte de Chaleur)

Définition Matrice colorée représentant l'intensité des valeurs.

Quand l'utiliser

- Visualiser des corrélations entre variables
- Montrer des patterns temporels (heure x jour)
- Analyser des matrices de confusion

Contexte Bancaire

Exemple: Matrice de corrélation des indicateurs de risque

- NPL vs ratio de couverture
- Score de crédit vs taux de défaut
- Identifie les variables redondantes

Interprétation

Échelle de couleur:

- Divergente (rouge-blanc-bleu): Pour corrélations (-1 à +1)
 - Séquentielle (blanc à bleu): Pour valeurs positives
 - Importance de la légende et des annotations
-

3.8 Area Chart (Graphique d'Aire)

Définition Similaire au line chart mais avec l'aire sous la courbe remplie.

Types

- **Simple:** Une série
- **Empilé:** Plusieurs séries cumulées (montre la composition)
- **Empilé 100%:** Proportions relatives

Quand l'utiliser

- Montrer l'évolution d'un total et ses composantes
- Visualiser des volumes cumulés

Contexte Bancaire

Exemple: Évolution des dépôts par type

- Dépôts à vue
 - Dépôts à terme
 - Épargne
 - Total visible + composition
-

3.9 Violin Plot

Définition Combinaison d'un box plot et d'un density plot, montrant la distribution complète.

Quand l'utiliser

- Comparer des distributions avec plus de détails que le box plot
- Visualiser des distributions multimodales

Contexte Bancaire

Exemple: Distribution des transactions par jour de la semaine

- Révèle si certains jours ont des patterns bimodaux
 - Plus informatif que le box plot seul
-

3.10 Treemap

Définition Représentation hiérarchique par des rectangles proportionnels.

Quand l'utiliser

- Visualiser des hiérarchies avec des valeurs
- Montrer la composition avec beaucoup de catégories

Contexte Bancaire

Exemple: Portefeuille de prêts par secteur et sous-secteur

- Niveau 1: Secteur (Agriculture, Commerce, ...)
 - Niveau 2: Sous-secteur (Céréales, Élevage, ...)
 - Taille = montant, couleur = risque
-

4. Graphiques pour l'Analyse Bancaire

4.1 Waterfall Chart

Usage Montre comment une valeur initiale est affectée par des additions et soustractions.

Contexte

Exemple: Évolution du résultat net
Revenus d'intérêts
+ Commissions
- Charges d'intérêts
- Provisions pour pertes
- Charges d'exploitation
= Résultat net

4.2 Funnel Chart

Usage Visualise les étapes d'un processus avec réduction progressive.

Contexte

Exemple: Processus d'octroi de crédit
Demandes reçues: 1000
↓ Pré-qualification: 700
↓ Analyse de crédit: 400
↓ Approbation: 250
↓ Décaissement: 200

4.3 Bullet Chart

Usage Compare une mesure à une cible avec des zones de performance.

Contexte

Exemple: Performance des KPIs
- Actuel vs objectif
- Zones: Mauvais / Moyen / Bon

4.4 Sparklines

Usage Mini-graphiques intégrés dans des tableaux ou dashboards.

Contexte

Exemple: Dashboard exécutif

Agence	Dépôts	Tendance
PAP	10M	
CAP	5M	

5. Choix du Graphique: Guide Décisionnel

5.1 Par Question Business

Question	Graphique Recommandé
"Comment se répartit le portefeuille?"	Pie, Treemap, Stacked bar
"Quelle est la tendance?"	Line chart, Area chart
"Quelles agences performant le mieux?"	Bar chart (trié)
"Y a-t-il une relation entre X et Y?"	Scatter plot
"Quelle est la distribution?"	Histogramme, Box plot
"Comment les groupes se comparent?"	Grouped bar, Box plot
"Où sont les anomalies?"	Box plot, Scatter plot

5.2 Par Type de Données

Variable catégorielle seule → Bar chart

Variable continue seule → Histogramme, Box plot

Catégorielle + Continue → Box plot par groupe

Continue + Continue → Scatter plot

Temporelle + Continue → Line chart

Hiérarchique → Treemap

6. Couleurs et Design

6.1 Palettes de Couleurs

Séquentielle Pour données ordonnées (faible à élevé)

Exemple: Intensité du risque

Bleu clair → Bleu foncé

Divergente Pour données avec point central significatif

Exemple: Performance vs objectif

Rouge (sous) → Blanc (=) → Vert (sur)

Qualitative Pour catégories distinctes sans ordre

Exemple: Types de produits

Bleu, Orange, Vert, Violet (couleurs distinctes)

6.2 Accessibilité

- 8% des hommes sont daltoniens
 - Éviter rouge/vert seuls
 - Utiliser des patterns en plus des couleurs
 - Tester avec des simulateurs de daltonisme
-

7. Outils de Visualisation

7.1 Power BI

- Dashboards interactifs
- Forte intégration Microsoft
- DAX pour calculs complexes

7.2 Python (Matplotlib, Seaborn, Plotly)

- Personnalisation totale
- Reproductibilité
- Analyse exploratoire

7.3 Excel

- Accessible et familier
- Limité pour grands volumes
- Pivot charts

7.4 Tableau

- Leader du marché
 - Drag-and-drop intuitif
 - Coûteux
-

8. Bonnes Pratiques pour le Contexte Bancaire

8.1 Confidentialité

- Anonymiser les données clients
- Agréger pour éviter l'identification
- Suivre les politiques internes

8.2 Précision

- Afficher les sources de données
- Indiquer la date de fraîcheur
- Mentionner les exclusions

8.3 Standardisation

- Utiliser les mêmes couleurs pour les mêmes concepts
- Formats de dates consistants
- Définitions claires des métriques

8.4 Storytelling avec les Données

Structure recommandée:

1. Contexte: Pourquoi ce graphique?
2. Observation: Que voyons-nous?
3. Insight: Qu'est-ce que cela signifie?
4. Action: Que devons-nous faire?

9. Exemples Pratiques Bancaires

9.1 Dashboard de Risque de Crédit

Composants:

- KPI cards: NPL ratio, Coverage ratio, Provision expense
- Trend line: Évolution du NPL sur 12 mois
- Bar chart: NPL par secteur économique
- Heatmap: Matrice de transition des ratings
- Box plot: Distribution des PD par segment

9.2 Tableau de Bord Commercial

Composants:

- Funnel: Pipeline de prêts
- Map: Performance par région
- Line chart: Objectifs vs réalisations
- Bar chart: Top 10 agents de crédit
- Gauge: Taux d'atteinte des objectifs

9.3 Analyse de la Clientèle

Composants:

- Scatter plot: RFM segmentation
- Treemap: Répartition par segment
- Histogram: Distribution de l'ancienneté
- Line chart: Évolution du churn
- Cohort analysis: Rétention par cohorte

10. Questions d'Entretien Fréquentes

1. **Quel graphique utiliseriez-vous pour montrer la distribution des montants de prêts?** → Histogramme (avec box plot en complément)
2. **Comment visualiser la relation entre le score de crédit et le taux de défaut?** → Scatter plot ou bar chart si score catégorisé
3. **Pourquoi le pie chart n'est pas toujours approprié?** → Difficile de comparer des parts similaires, limité en nombre de catégories
4. **Qu'indique un box plot avec une médiane proche du Q1?** → Distribution asymétrique à droite (positive skew)
5. **Comment choisir le nombre de bins dans un histogramme?** → Règle de Sturges ou Freedman-Diaconis, itérer pour clarté

11. Checklist de Validation d'un Graphique

- ☐ Le titre est clair et informatif
- ☐ Les axes sont correctement labellisés avec unités

- ☐ La légende est présente et lisible
 - ☐ Les couleurs sont appropriées et accessibles
 - ☐ L'axe Y commence à zéro (pour bar charts)
 - ☐ La source des données est indiquée
 - ☐ Le graphique répond à une question business claire
 - ☐ Le message principal est immédiatement visible
-

Résumé

Graphique	Usage Principal	Éviter si
Histogramme	Distribution continue	Peu de données
Box plot	Comparaison de distributions	Audience non-technique
Bar chart	Comparaison de catégories	Trop de catégories
Line chart	Tendances temporelles	Données non ordonnées
Scatter plot	Relations entre variables	Variables catégorielles
Pie chart	Composition simple	> 5 catégories
Heatmap	Corrélations, patterns	Pas de pattern clair

Rappel final: Le meilleur graphique est celui qui communique le message le plus clairement à votre audience spécifique.