

Test Data Visualization - Test 2

Sujet: Visualisation de Données

Niveau: Intermédiaire

Nombre de questions: 25

Questions et Réponses

Q1. Expliquez le concept de “pre-attentive attributes” en visualisation.

R1. Les **attributs pré-attentifs** sont des propriétés visuelles que le cerveau traite inconsciemment et instantanément (<500ms): - **Couleur** (teinte, saturation) - **Forme** (orientation, taille) - **Position** (axe X/Y) - **Mouvement** (animation)

Application: Utiliser ces attributs pour mettre en évidence les données importantes (ex: rouge pour les alertes).

Q2. Quelle est la hiérarchie de perception visuelle pour les encodages quantitatifs?

R2. Du plus précis au moins précis (selon Cleveland & McGill): 1. **Position sur échelle commune** (le plus précis) 2. **Position sur échelles non-alignées** 3. **Longueur** 4. **Angle/Pente** 5. **Aire** 6. **Volume** 7. **Couleur/Saturation** (le moins précis)

Implication: Préférer les bar charts (position/longueur) aux pie charts (angle/aire).

Q3. Comment visualiseriez-vous des données avec une distribution très asymétrique (skewed)?

R3. Options: 1. **Transformation log:** Échelle logarithmique sur l'axe Y 2. **Box plot:** Montre médiane et quartiles (robuste aux outliers) 3. **Histogramme avec bins adaptés:** Bins plus étroits dans la zone dense 4. **Violin plot:** Montre la forme asymétrique

En Python

```
plt.hist(data, bins=50)
plt.xscale('log') # ou
plt.yscale('log')
```

Q4. Qu'est-ce que le “lie factor” et comment l'éviter?

R4. Le **Lie Factor** (Tufte) mesure la distorsion visuelle:

Lie Factor = (Taille de l'effet visuel) / (Taille de l'effet dans les données)

Objectif: Lie Factor = 1 (représentation fidèle)

Sources de distorsion à éviter: - Axes tronqués non signalés - Effets 3D - Aires disproportionnées - Échelles non-linéaires non indiquées

Q5. Comment représenter l'incertitude dans une visualisation?

R5. Méthodes: 1. **Barres d'erreur:** \pm écart-type ou intervalle de confiance 2. **Bandes de confiance:** Zones ombrées autour d'une ligne 3. **Box plots:** Montrent la variabilité 4.

Distributions complètes: Violin plots, histogrammes 5. **Annotations textuelles:** "IC 95%: [X, Y]"

Q6. Quand utiliser une échelle logarithmique?

R6. Utiliser l'échelle log quand: - Les données couvrent **plusieurs ordres de grandeur** - On s'intéresse aux **taux de croissance** (pas aux différences absolues) - La distribution est **log-normale** - On compare des **ratios**

Exemple bancaire: Montants de prêts (1K à 100M HTG)

Q7. Comment éviter le "chart junk" (pollution visuelle)?

R7. Éliminer: - **Effets 3D** sans justification - **Grilles denses** - **Bordures lourdes** - **Couleurs de fond** - **Légendes redondantes** (si labels directs possibles) - **Décorations** (clipart, images)

Règle: Chaque élément doit servir la compréhension des données.

Q8. Comment concevoir un dashboard pour différents niveaux hiérarchiques?

R8. | Niveau | Focus | Visualisation | |----|----|-----| | **CEO/Board** | KPIs agrégés, tendances | Scorecards, sparklines | | **Direction** | Performance par segment | Graphiques comparatifs | | **Managers** | Détails opérationnels | Tableaux détaillés, filtres | | **Analystes** | Exploration | Dashboards interactifs |

Principe: Drill-down du général au détaillé.

Q9. Quelle est la règle pour le nombre de couleurs dans un graphique?

R9. - **Maximum 5-7 couleurs** distinctes - Au-delà, utiliser des **variations de teinte** (clair/foncé) - **Couleur de mise en évidence:** Une couleur vive + reste en gris - **Séquence:** Gradient pour données ordonnées - **Divergente:** Deux couleurs opposées pour comparaison (ex: positif/négatif)

Q10. Comment gérer les valeurs manquantes dans une visualisation?

R10. Options selon le contexte: 1. **Line chart:** Interrompre la ligne ou utiliser pointillés 2. **Bar chart:** Laisser un espace avec annotation "N/A" 3. **Heatmap:** Couleur distincte pour "missing" 4. **Annotation:** Indiquer clairement le pourcentage de données manquantes

Ne jamais: Afficher zéro à la place de manquant.

Q11. Comment visualiser des données géographiques pour les agences bancaires?

R11. - **Carte choroplète:** Régions colorées selon une métrique - **Carte à bulles:** Taille proportionnelle à une valeur - **Carte de chaleur géographique:** Densité de clients/transactions

Outils: Power BI Maps, Folium (Python), Tableau

Q12. Qu'est-ce qu'un small multiple et quand l'utiliser?

R12. Les **small multiples** sont plusieurs petits graphiques identiques, chacun montrant un sous-ensemble des données: - Même échelle - Même type de graphique - Différente catégorie

Utilisation: Comparer les tendances de plusieurs agences, produits, ou segments.

Q13. Comment rendre un graphique accessible (daltonisme, etc.)?

R13. - **Éviter rouge-vert** ensemble (8% des hommes daltoniens) - Utiliser des **patterns** en plus de la couleur - **Labels directs** sur les éléments - **Contraste suffisant** (WCAG 4.5:1) - **Taille de police** lisible (min 10-12pt) - **Alt text** pour les versions numériques

Q14. Quelle visualisation pour comparer actual vs budget vs forecast?

R14. **Combo chart** avec: - Barres: Actual - Ligne continue: Budget - Ligne pointillée: Forecast
Ou **bar chart groupé** si les valeurs sont à des dates spécifiques.

Q15. Comment visualiser un grand nombre de catégories (>20)?

R15. Stratégies: 1. **Regrouper:** Top 10 + "Autres" 2. **Treemap:** Montre toutes les catégories hiérarchiquement 3. **Bar chart horizontal:** Scrollable ou paginé 4. **Filtre interactif:** L'utilisateur sélectionne les catégories 5. **Small multiples:** Si comparaison temporelle

Q16. Qu'est-ce qu'un dual-axis chart et quels sont ses risques?

R16. Un graphique à **double axe Y** pour deux métriques différentes.

Risques: - Échelles arbitraires → fausses corrélations - Confusion pour le lecteur - Manipulation possible des échelles

Règle: Utiliser uniquement si les métriques sont liées conceptuellement ET avec des échelles justifiées.

Q17. Comment intégrer des données textuelles dans une visualisation?

R17. Options: - **Word cloud:** Fréquence des mots (commentaires clients) - **Sentiment timeline:** Score sentiment dans le temps - **Bar chart:** Fréquence des thèmes/catégories - **Annotations:** Extraits clés sur un graphique temporel

Q18. Quelle est la différence entre un dashboard opérationnel et analytique?

R18. | Opérationnel | Analytique | -----|-----| | Temps réel ou quotidien | Historique, tendances | | Actions immédiates | Insights, patterns | | KPIs de monitoring | Analyse causale | | Alertes, seuils | Exploration, drill-down | | Utilisateur: Opérations | Utilisateur: Management |

Q19. Comment représenter des intervalles de temps (durées)?

R19. - **Gantt chart:** Périodes de début à fin - **Timeline:** Événements séquentiels - **Histogram:** Distribution des durées - **Box plot:** Comparer durées entre groupes

Exemple bancaire: Durée de traitement des demandes de prêt.

Q20. Quels sont les principes de Gestalt appliqués à la visualisation?

R20. Principes de perception: 1. **Proximité:** Éléments proches semblent groupés 2. **Similitude:** Éléments similaires (couleur, forme) semblent liés 3. **Continuité:** L'œil suit les lignes et courbes 4. **Fermeture:** Le cerveau complète les formes incomplètes 5. **Figure-fond:** Distinction entre sujet et arrière-plan

Application: Utiliser l'espace et les couleurs pour créer des groupements logiques.

Q21. Comment créer un graphique “storytelling”?

R21. Éléments: 1. **Titre accrocheur:** Message principal 2. **Contexte:** Pourquoi c'est important 3. **Point focal:** Mise en évidence du message clé 4. **Annotations:** Explications des points importants 5. **Call to action:** Que faire avec cette information

Exemple: “Les défauts agricoles ont doublé depuis janvier” + annotation sur les causes + recommandation.

Q22. Comment visualiser des données de séries temporelles avec saisonnalité?

R22. - **Décomposition:** Trend + Seasonal + Residual - **Year-over-year overlay:** Plusieurs années superposées - **Heatmap calendrier:** Jours x mois avec intensité - **Seasonal sub-series:** Un graphique par saison/mois

```
from statsmodels.tsa.seasonal import seasonal_decompose  
result = seasonal_decompose(data, model='additive')  
result.plot()
```

Q23. Quelle visualisation pour montrer des données en temps réel?

R23. - **Gauges/Speedometers:** Valeur actuelle vs seuil - **Live line chart:** Mise à jour continue - **Status indicators:** Vert/Jaune/Rouge - **Ticker/Feed:** Valeurs défilantes

Contexte bancaire: Monitoring des transactions, alertes fraude.

Q24. Comment documenter et standardiser les visualisations dans une organisation?

R24. Créer un **Style Guide** incluant: - Palette de couleurs officielle - Typographie - Types de graphiques approuvés - Templates Power BI/Excel - Règles de labellisation - Exemples de bonnes/mauvaises pratiques

Q25. Concevez un mini-dashboard pour le suivi du risque crédit (4 visualisations max).

R25.

1. NPL RATIO [Bullet Chart] Actual: 4.8% Target: 5.0%	2. MIGRATION BUCKETS [Stacked Bar] Current 30j 60j 90j+
--	---

3. NPL PAR SECTEUR [Bar Chart Horizontal] Agriculture: 8%	4. TENDANCE 12 MOIS [Line Chart] NPL vs Coverage Ratio
---	--

Commerce: 5%

Services: 3%

Principes appliqués: - Max 4 graphiques pour focus - KPI principal en haut à gauche - Cohérence des couleurs - Actions claires par graphique

Scoring

Score	Niveau
0-10	À améliorer
11-17	Intermédiaire
18-22	Avancé
23-25	Expert