

Représentation des données

Checklist for good graphics

Jean-Marc.Vincent@imag.fr



Grenoble 2017
MIAGE

PRINCIPES POUR LA VISUALISATION

1 **PROBLÈME : Représentation des données**

2 **PRINCIPES : Idées de base**

3 **Common Mistakes**

4 **PRINCIPES : indicateurs**

5 **EXEMPLES : illustration**

6 **SYNTHÈSE**

LA PRÉSENTATION

Problème : fournir de jolies images pour aider à la compréhension

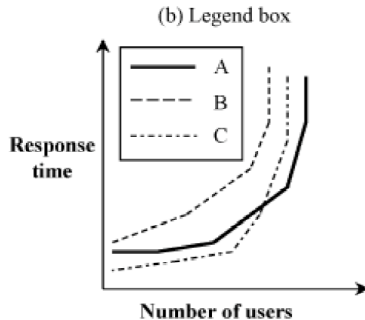
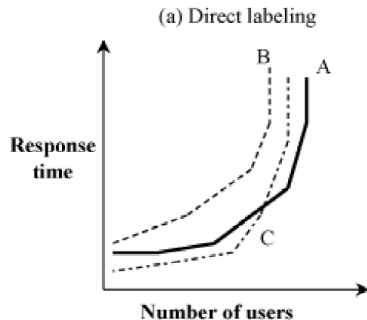
- ▶ **Améliorer grandement la qualité d'un article**
- ▶ Montrer la qualité de votre travail
- ▶ Les images/figures/diagrammes entraînent des discussions
- ▶ et génèrent de nouveaux problèmes

Erreurs de présentation

- ▶ **Sémantique des objets graphiques**
- ▶ conventions pour les schémas/diagrammes...
- ▶ premiers pas vers la validation

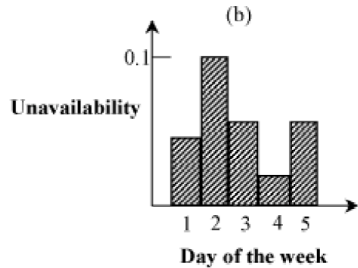
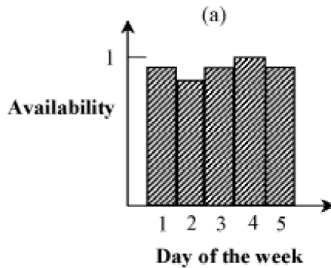
GUIDELINES FOR GOOD GRAPHICS (JAIN)

Minimum effort for the reader



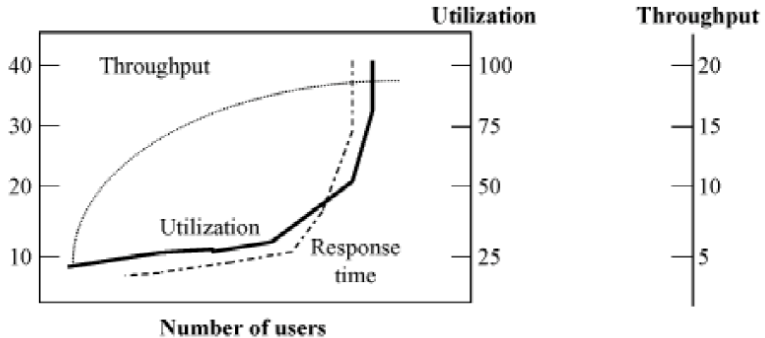
GUIDELINES FOR GOOD GRAPHICS (JAIN)

Minimize Ink



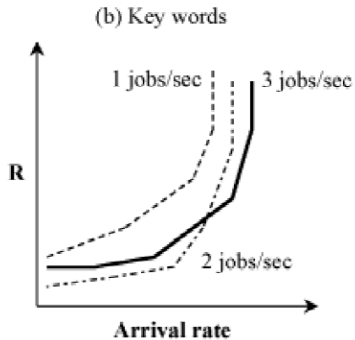
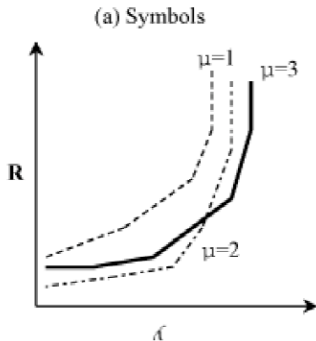
COMMON MISTAKES

Multiple scaling, Too much information



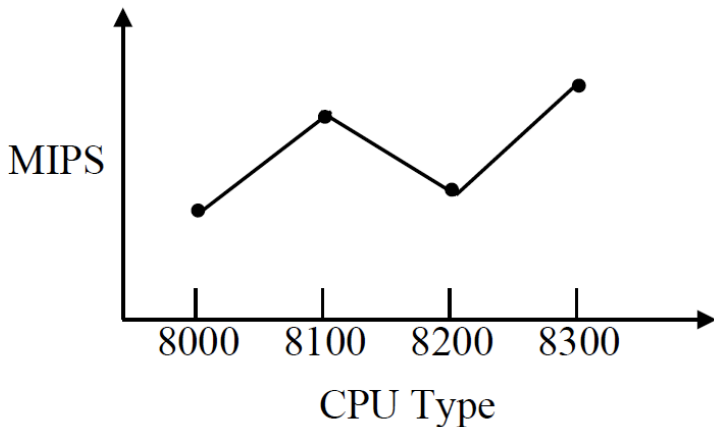
COMMON MISTAKES

Cryptic information



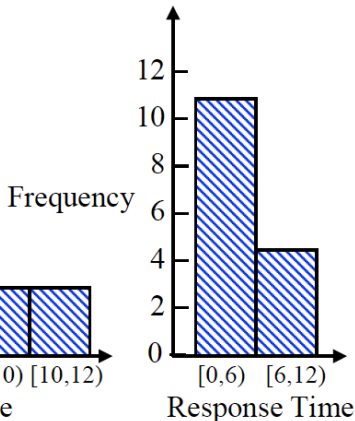
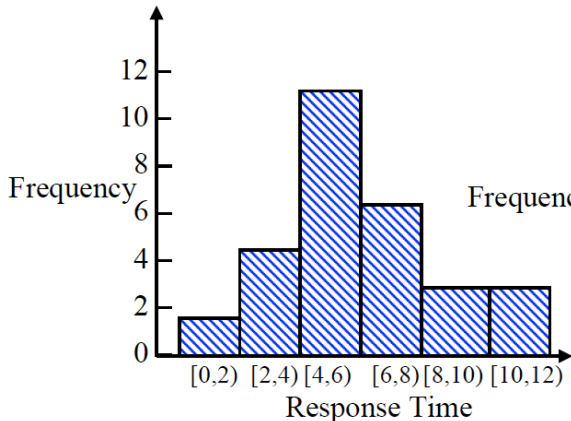
COMMON MISTAKES

Non-relevant graphic objects



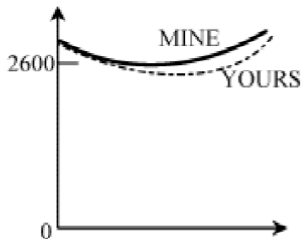
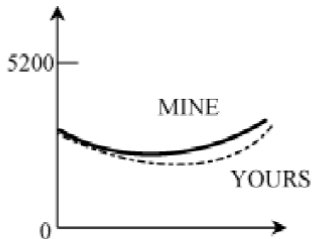
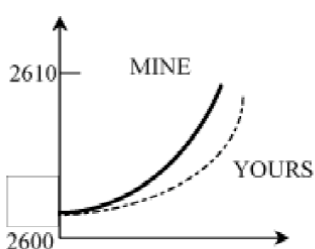
COMMON MISTAKES

Non-relevant graphic objects



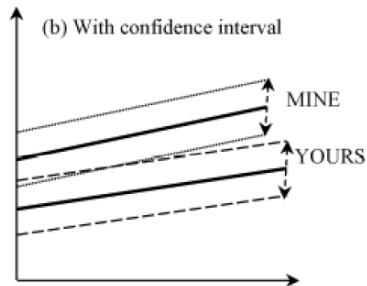
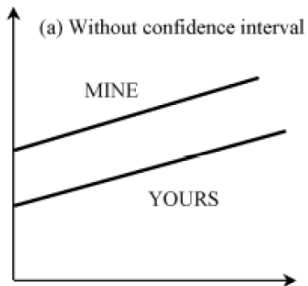
COMMON MISTAKES

Howto cheat ?



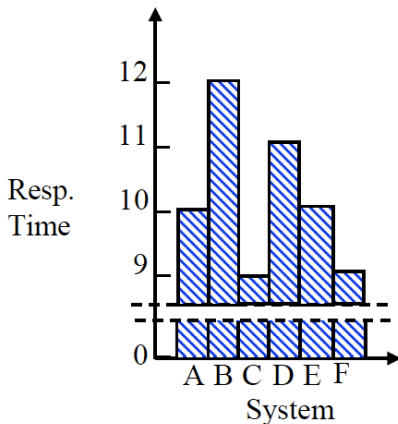
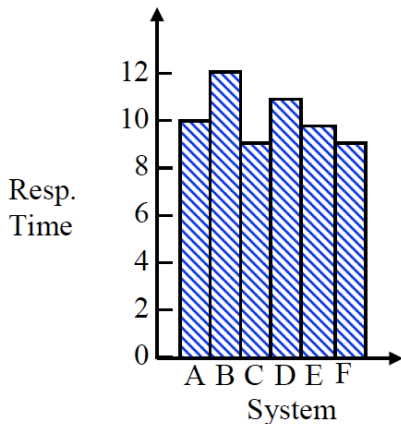
COMMON MISTAKES

Howto cheat ?



COMMON MISTAKES

Howto cheat ?



DONNÉES

Compréhension des données visualisées

- ▶ Le type de graphique est adapté aux données (courbe, bâtons, histogramme, nuages,...) ;
- ▶ Les interpolations/approximations ont un sens ;
- ▶ Les courbes sont définies avec suffisamment de points ;
- ▶ La méthode de construction de la courbe est claire : interpolation (linéaire, polynomiale,...), approximation (polynomiale, régression,...)
- ▶ Les intervalles de confiance sont visualisés (ou donnés séparément) ;
- ▶ Le pas de l'histogramme est adapté ;
- ▶ L'ordonnée de l'histogramme est la probabilité (de 0 à 1).

OBJETS GRAPHIQUES

Éléments graphiques

- ▶ Les objets graphiques sont lisibles sur écran, impression couleur/noir et blanc, vidéo...
- ▶ La palette de couleur est standard, sans couleurs proches, sans vert (video) ;
- ▶ Les axes du graphique sont bien identifiés et annotés ;
- ▶ Les échelles et les unités sont marquées ;
- ▶ Les courbes se croisent sans ambiguïté, en un point repérable ;
- ▶ Les grilles aident à lire la courbe.

ANNOTATIONS

Annotations et commentaires

- ▶ Les axes sont bien annotés par les quantités ;
- ▶ Le nom des axes est clair, concis et suffisamment explicatif ;
- ▶ Les unités sont indiquées sur les axes ;
- ▶ Les axes sont orientés du plus petit au plus grand de gauche à droite et de bas en haut ;
- ▶ L'origine est (0, 0) sinon c'est justifié ;
- ▶ Les échelles sont contigües.
- ▶ Pour les diagrammes en bâton/histogrammes l'ordre des bâtons est systématique (alphabétique, temporel, du meilleur au moins bon sont préférables à un ordre aléatoire)
- ▶ Les courbes sont annotées individuellement (légende) ;
- ▶ Les barres/rectangles sont annotés individuellement (légende).

MESSAGE

Information : ce que dit le graphique

- ▶ Les courbes sont mises sur la même échelle ;
- ▶ Le nombre de courbes est petit (pas plus de 6) ;
- ▶ Les courbes à comparer sont sur le même graphique ;
- ▶ On ne peut pas supprimer une courbe sans diminuer significativement l'information donnée par le graphique ;
- ▶ Le graphique fournit une information pertinente au lecteur ;
- ▶ Si l'axe vertical représente une quantité aléatoire, il présente des barres d'erreur ;
- ▶ Peut-on supprimer des objets graphiques, du texte, des courbes sans modifier la compréhension ?

CONTEXTE

Contexte du graphique

- ▶ Tous les symboles du graphique sont référencés (définis) dans le texte ;
- ▶ Les variables tracées fournissent plus d'informations que des représentations alternatives ; (choix d'autres variables)
- ▶ Le graphique a un titre ;
- ▶ Le titre est autosuffisant pour comprendre partiellement le graphique ;
- ▶ La figure est référencée dans le corps du texte ;
- ▶ La figure est commentée dans le texte.

EXAMPLE (1)

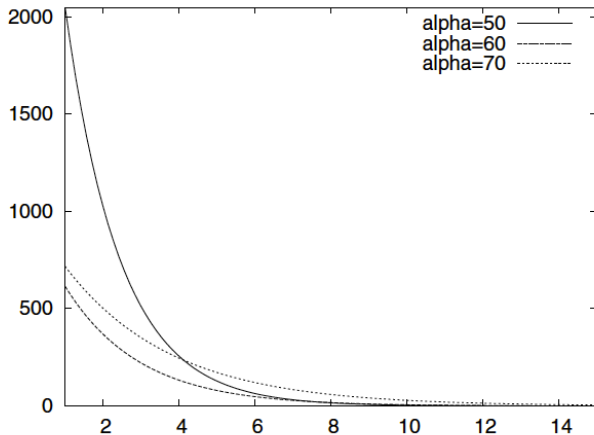


Figure 1. Temperature cooling scheme for different α values

EXAMPLE (2)

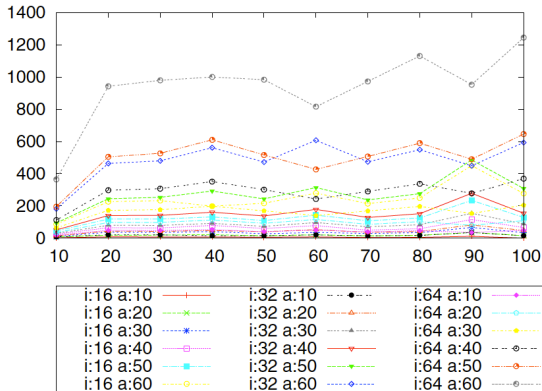


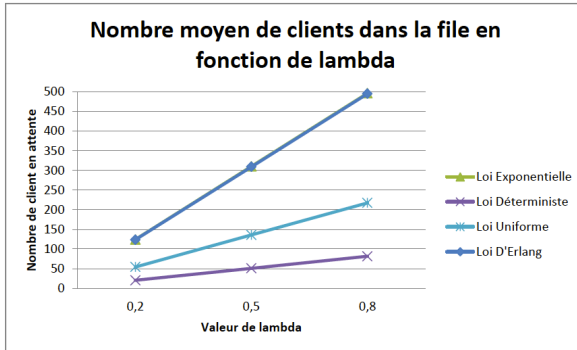
Figure 2. SA – 512-node grid – pdf net 10 – i
stands for iterations and a for α

EXAMPLE (3)

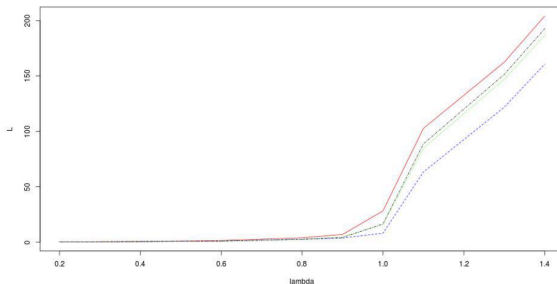
Table 7. Response Time – Grid 512 – 64 processes – pdf net 1000

#Apps	Lowest	SA	Route	Round Robin
	MRT-Stdev	MRT-Stdev	MRT-Stdev	MRT-Stdev
10	113.75 \pm 8.65E3	62.89 \pm 1.4E3	110.39 \pm 1.01E4	129.62 \pm 10890.67
20	212.71 \pm 1.66E4	68.40 \pm 1.2E3	205.04 \pm 1.66E4	226.10 \pm 19521.60
30	324.89 \pm 7.33E4	71.50 \pm 2.5E3	213.69 \pm 2.41E4	262.34 \pm 17499.49
40	173.35 \pm 5.75E3	52.53 \pm 5.7E2	193.34 \pm 1.71E4	198.10 \pm 19750.03
50	483.98 \pm 1.25E5	109.16 \pm 5.2E3	252.77 \pm 3.23E4	390.86 \pm 129192.90
60	510.11 \pm 6.04E4	305.23 \pm 1.2E4	369.09 \pm 2.43E4	505.03 \pm 50711.91
70	716.14 \pm 4.95E5	508.84 \pm 1.5E5	529.45 \pm 2.71E5	771.57 \pm 307970.01
80	1235.14 \pm 3.02E6	827.67 \pm 2.4E6	1002.22 \pm 3.08E6	1122.46 \pm 2648874.68
90	589.96 \pm 8.68E4	404.34 \pm 2.9E4	446.83 \pm 7.48E4	663.13 \pm 112604.98
100	517.51 \pm 1.00E5	350.85 \pm 1.0E4	362.64 \pm 3.44E4	599.41 \pm 167523.40

EXAMPLE (4)



EXAMPLE (5)



Légende

Rouge : loi exponentielle

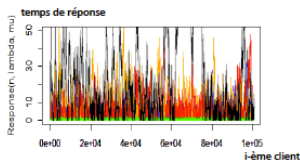
Bleu : loi uniforme comprise entre 0 et 2

Vert : loi uniforme entre 0,5 et 1,5

Noir : loi d'Erlang.

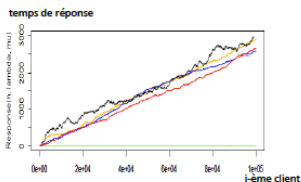
EXAMPLE (6)

Nous lançons à présent une simulation pour chaque loi avec un taux d'arrivée moyen $\lambda=1,9$ et un taux de service moyen $\mu=2$. Le régime est stable : ($\lambda < \mu$) :



L'ensemble des temps de réponse est borné par 50 secondes contrairement à 250 précédemment.

Nous lançons à présent une simulation pour chaque loi avec un taux d'arrivée moyen $\lambda=2$ et un taux de service moyen $\mu=1,9$. C'est la saturation ($\lambda > \mu$). :

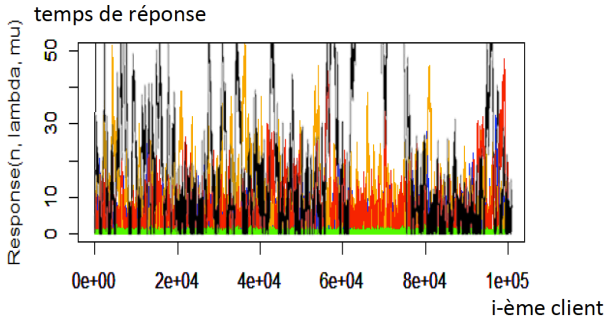


Nous constatons que les trajectoires obtenues sont quasiment linéaires. Cela est causé par le nombre supérieur de clients arrivant par unité de temps par rapport au nombre de clients pouvant être traité par le serveur.

2. Conclusion

Nous pouvons en conclure qu'il convient de privilégier de régime stationnaire au régime transitoire lors de l'analyse de trajectoires. En effet, le nombre de paramètres dont dépend la génération faite par le simulateur en régime transitoire rend l'analyse très difficile. Cependant, pour ce régime ainsi que le régime stationnaire, nous pouvons affirmer que pour les lois précédemment utilisées, les trajectoires sont composées de cycles et il existe des points de régénérations.

EXAMPLE (7)



Temps moyen de réponse en fonction de lambda
Régime stationnaire avec $\lambda = 1,9$ et $\mu = 2$ ($\lambda < \mu$)

EXAMPLE (8)

Screen-shots

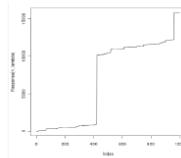


Figure 1 - Loi exponentielle

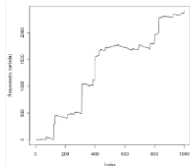


Figure 2 - loi uniforme sur l'intervalle [0,2]

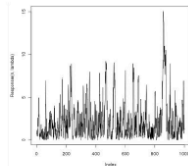
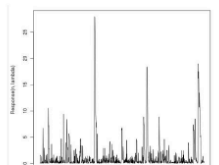
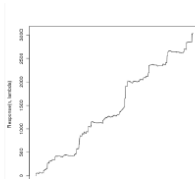


Figure 3 - loi uniforme sur l'intervalle [0.5;1.5]



EXAMPLE (9)

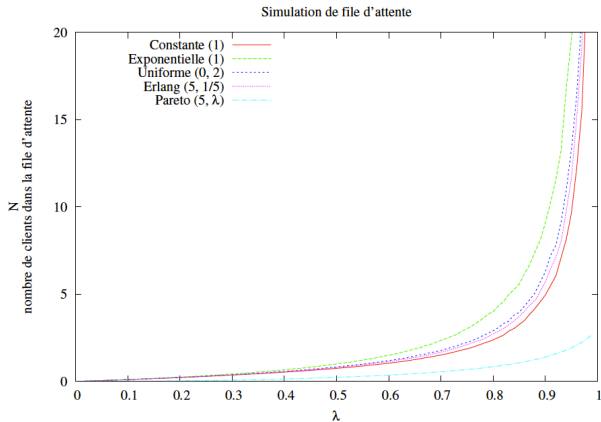


FIGURE 1 – Impact de la variabilité du temps de service sur le nombre moyen de clients dans la file d'attente

EXAMPLE (10)

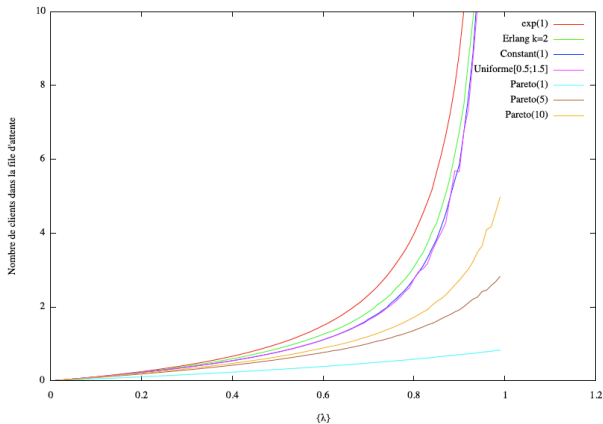


Figure : Evolution de la file d'attente en fonction du débit (arrivée et traitement)

SYNTHÈSE

Indications pour la conception d'une représentation graphique de données

- ▶ Demander un effort minimal au lecteur ;
- ▶ Maximiser l'information ;
- ▶ Minimiser l'*encre* ;
- ▶ Utiliser les conventions habituelles
- ▶ Faire plusieurs représentations avant de choisir la meilleure.

Erreurs classiques à éviter

- ▶ Trop d'objets graphiques
- ▶ Confusion d'échelles
- ▶ Symboles à la place de texte
- ▶ Information hors de propos
- ▶ Mauvais choix d'échelle/plages de valeurs

SYNTHÈSE

Toujours avoir à l'esprit : Qui doit lire/interpréter le graphique et pourquoi ?

Principes

Rasoir d'Occam Si deux représentations graphiques ont le même pouvoir d'information, choisir la représentation la plus simple.

Principe de complétude (Dijkstra) Quand on ne peut enlever un simple objet de la représentation graphique, alors elle est complète.

Principe de bon sens Utiliser le niveau de sophistication adapté.

Adapté des livres

- 1 Raj Jain. *The Art of Computer Systems Performance Analysis : Techniques for Experimental Design, Measurement, Simulation, and Modeling*. John Wiley & Sons, 1991.
- 2 Jean-Yves Le Boudec. *Performance Evaluation of Computer and Communication Systems*. EPFL Press, Lausanne, Switzerland, 2010.

Enfin : Le graphique doit être beau