

# What a nice picture !

## Data Visualization, an introduction

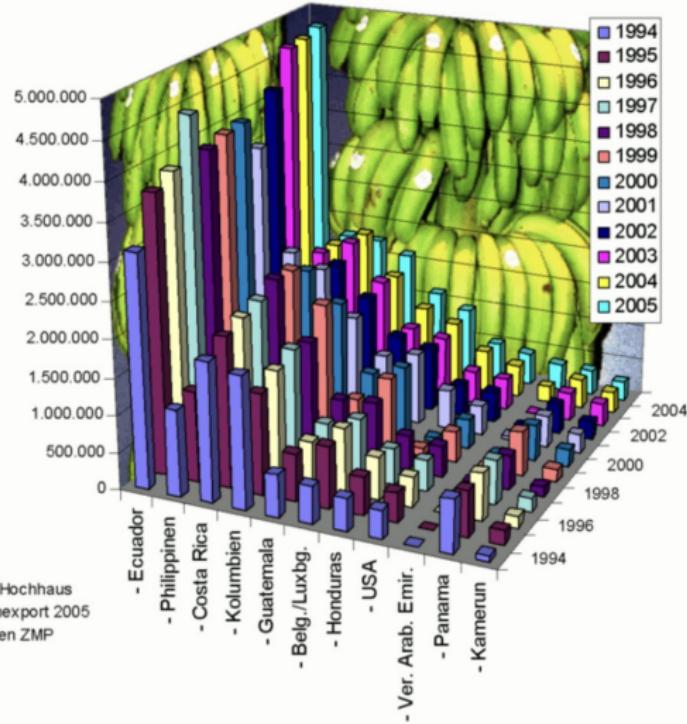
{Elise.Arnaud,Arnaud.Legrand,Jean-Marc.Vincent}@univ-grenoble-alpes.fr

Université de Grenoble-Alpes, UFR IM<sup>2</sup>AG  
M2R-MOSIG Scientific Methodology



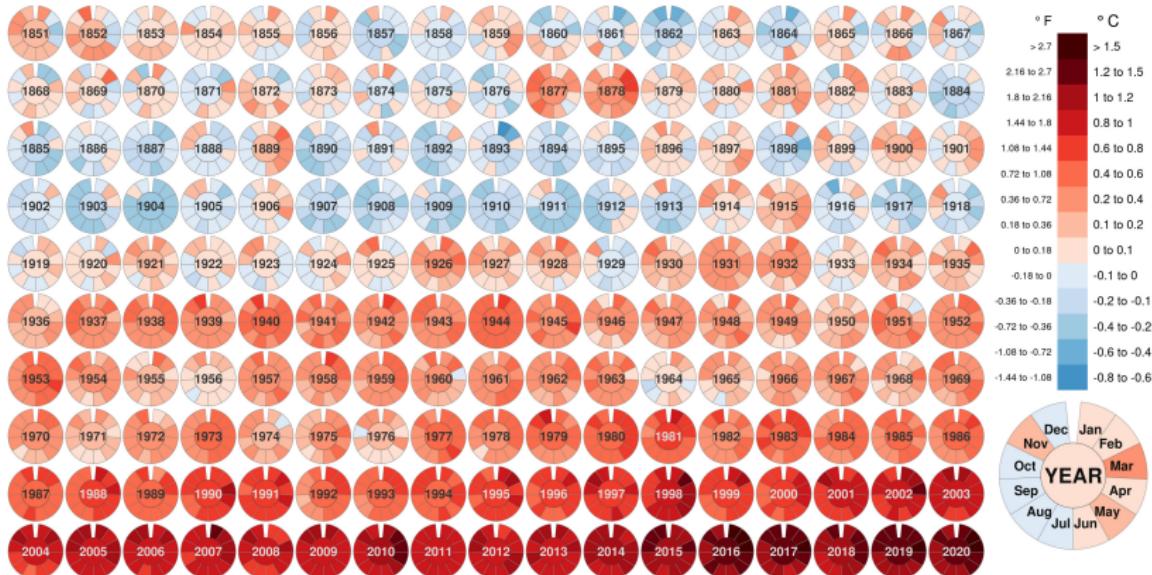
October 2021

### Export von Bananen in Tonnen von 1994-2005



Dr. Hochhaus  
Banlexport 2005  
Daten ZMP

## Monthly global mean temperature 1851 to 2020 (compared to 1850-1900 averages)

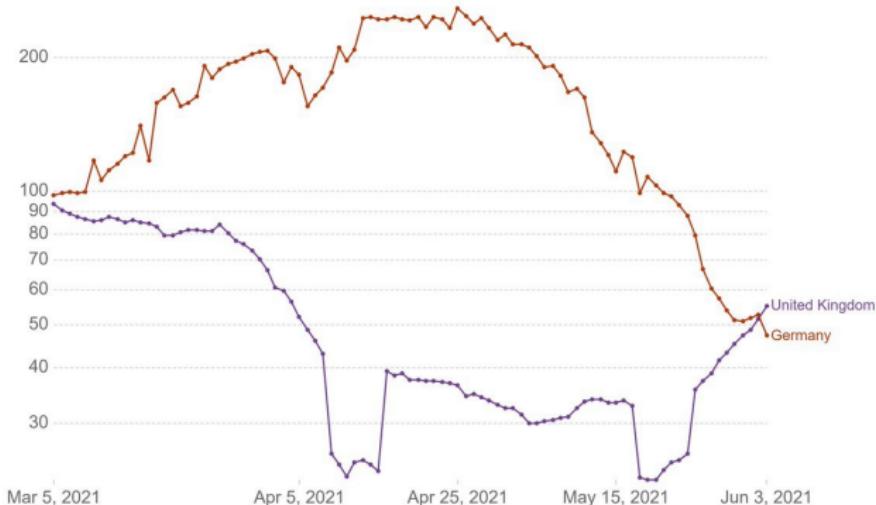


Data: HadCRUT5 - Created by: @neilrkaye

## Daily new confirmed COVID-19 cases per million people

Shown is the rolling 7-day average. The number of confirmed cases is lower than the number of actual cases; the main reason for that is limited testing.

Our World  
in Data

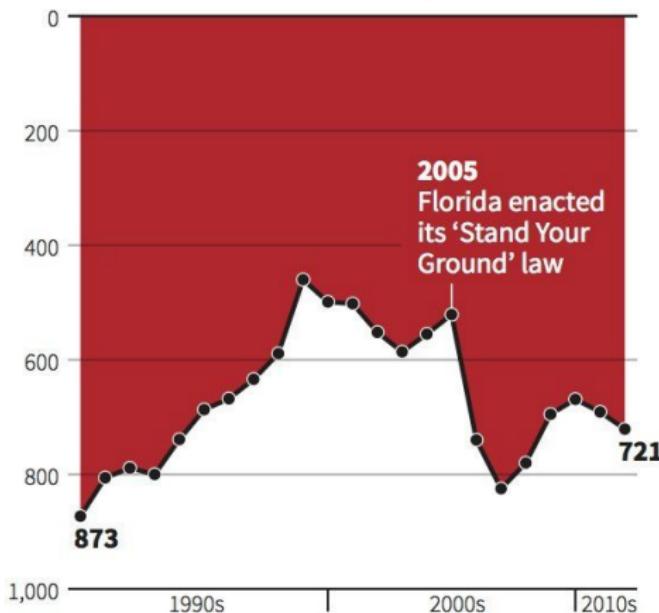


Source: Johns Hopkins University CSSE COVID-19 Data

CC BY

# Gun deaths in Florida

Number of murders committed using firearms



Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

REUTERS

# A GRAPHIC : A POWERFUL COMMUNICATION TOOL

## Motivation for Graphics

- ▶ Synthesis of the information
- ▶ Explore datasets
- ▶ Visual tests
- ▶ Communication of results

## Criteria for good graphics

- ▶ Readability for the reader
- ▶ Intelligibility of the message to the reader
- ▶ No possible misunderstanding

# HOW TO REPORT RESULTS FROM MODEL ANALYSIS

**Problem : provide "nice" pictures to help the understanding**

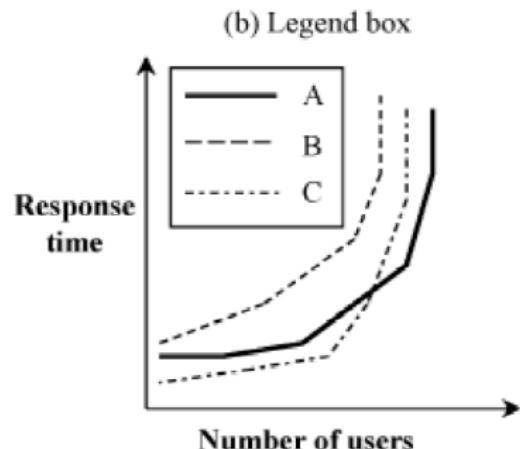
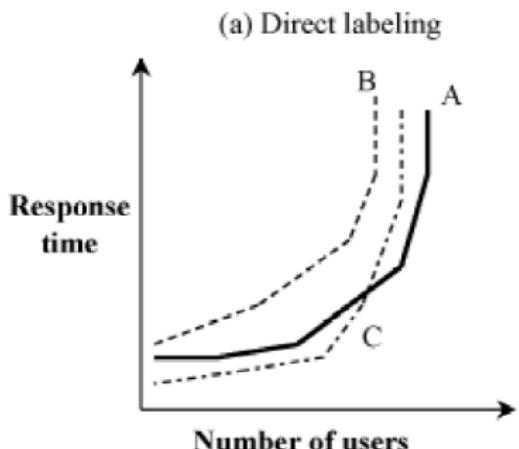
- ▶ Increases deeply the quality of a paper
- ▶ Show the scientific quality of your research
- ▶ Observation leads to open problems
- ▶ Pictures generates discussions

## Mistakes

- ▶ semantic of graphical objects
- ▶ conventions for graphics reading
- ▶ first step in scientific validation

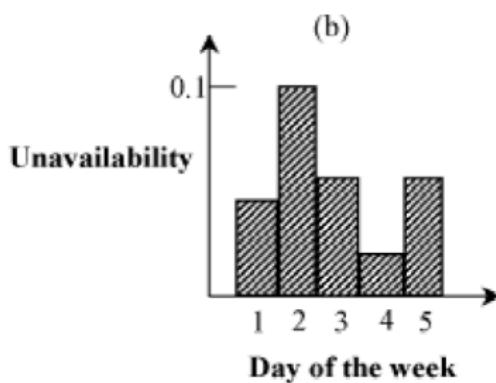
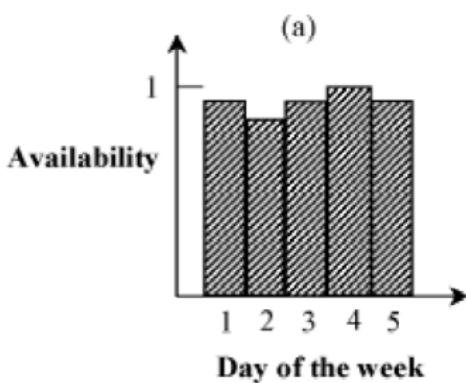
# GUIDELINES FOR GOOD GRAPHICS (JAIN)

Minimum effort for the reader



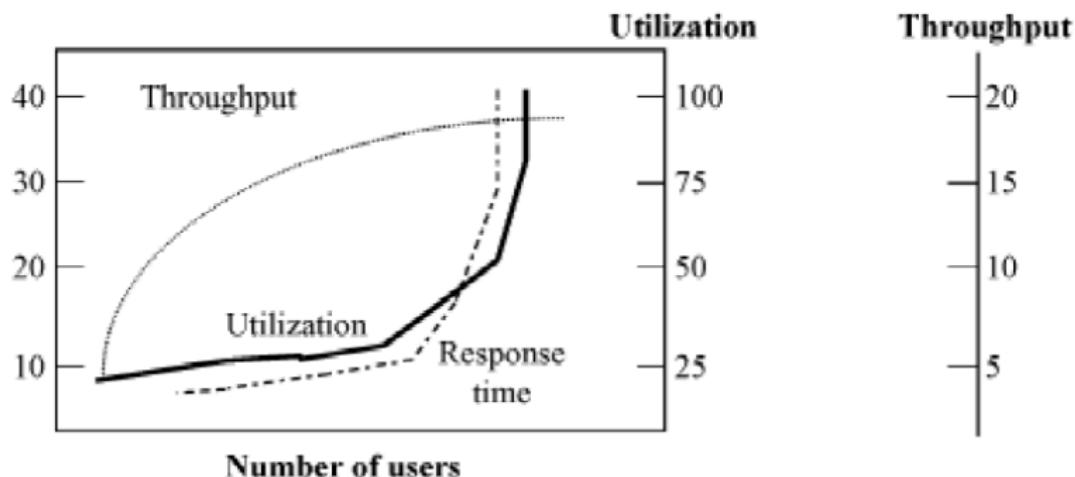
# GUIDELINES FOR GOOD GRAPHICS (JAIN)

## Minimize Ink



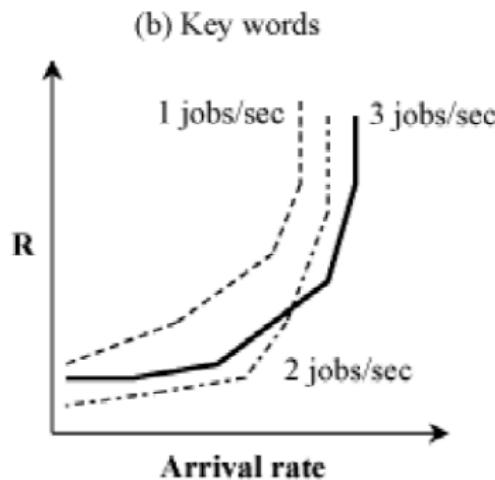
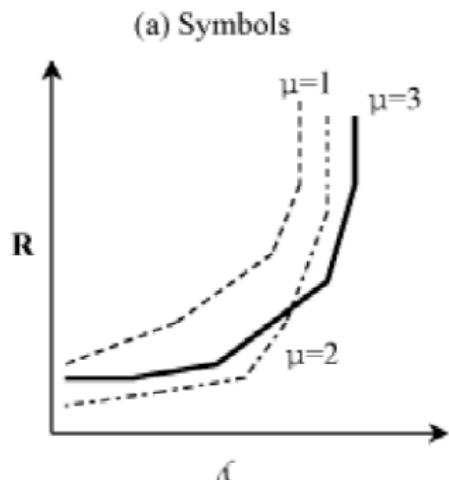
# COMMON MISTAKES

Multiple scaling, Too much information



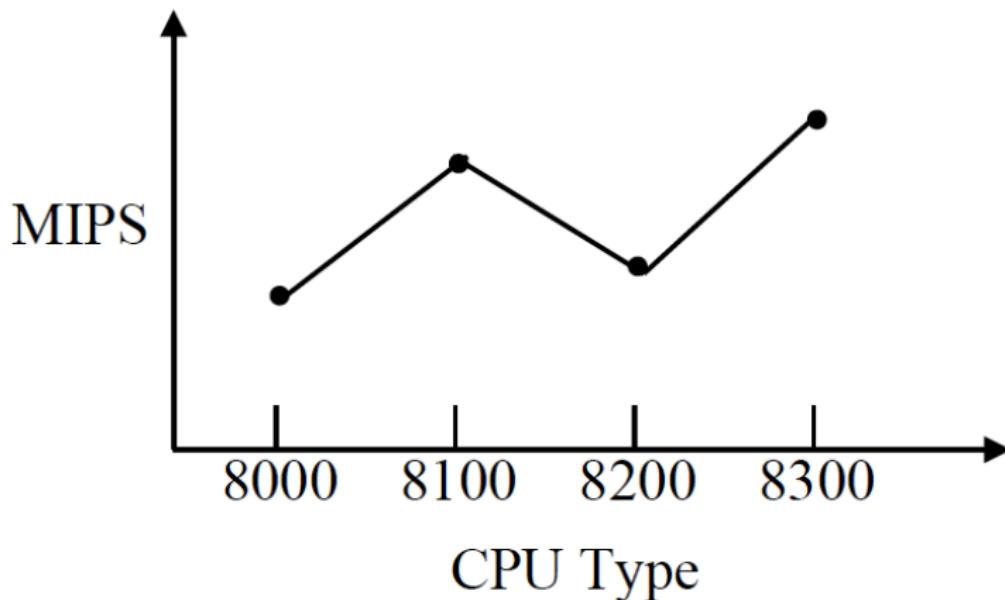
# COMMON MISTAKES

## Cryptic information



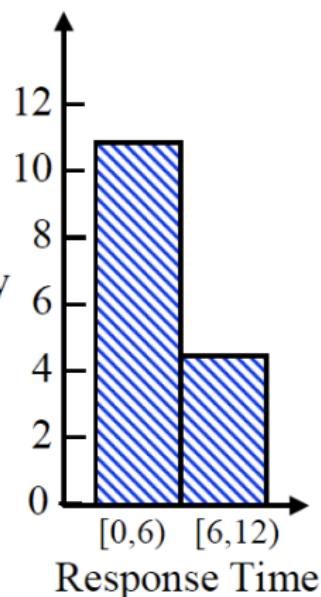
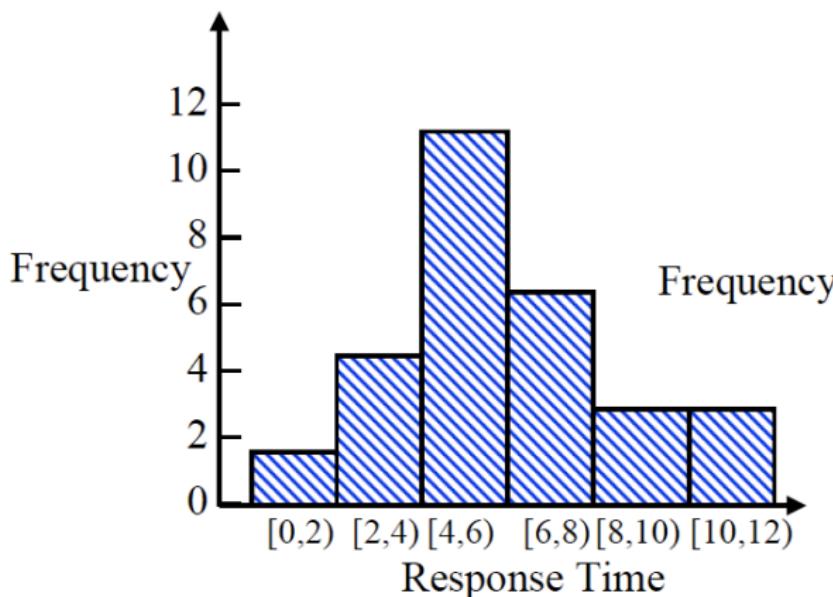
## COMMON MISTAKES

**Non-relevant graphic objects**



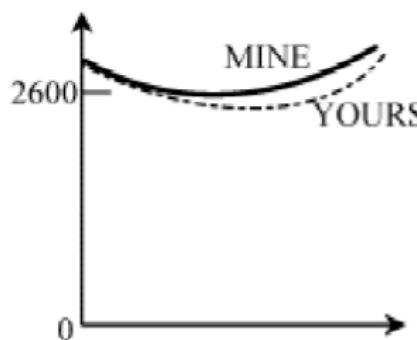
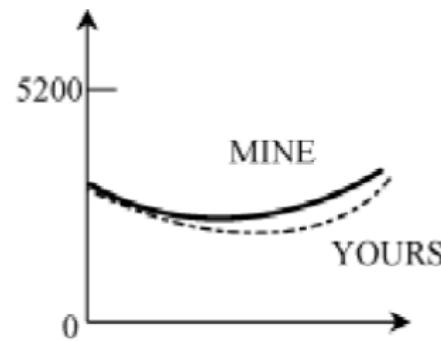
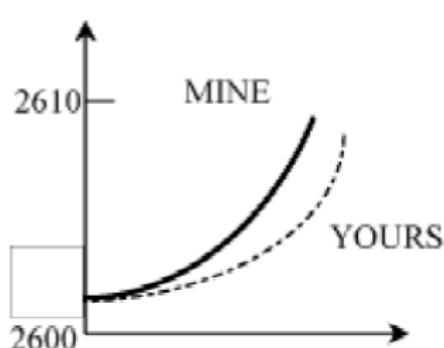
# COMMON MISTAKES

## Non-relevant graphic objects



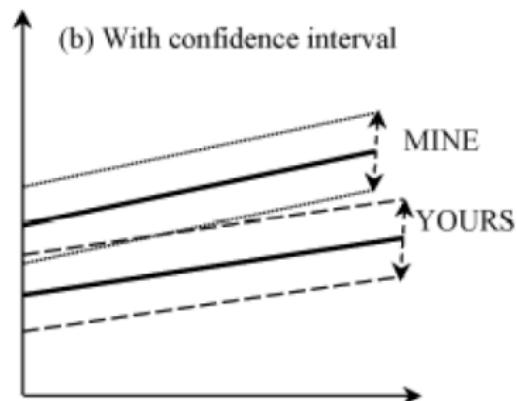
# COMMON MISTAKES

Howto cheat ?



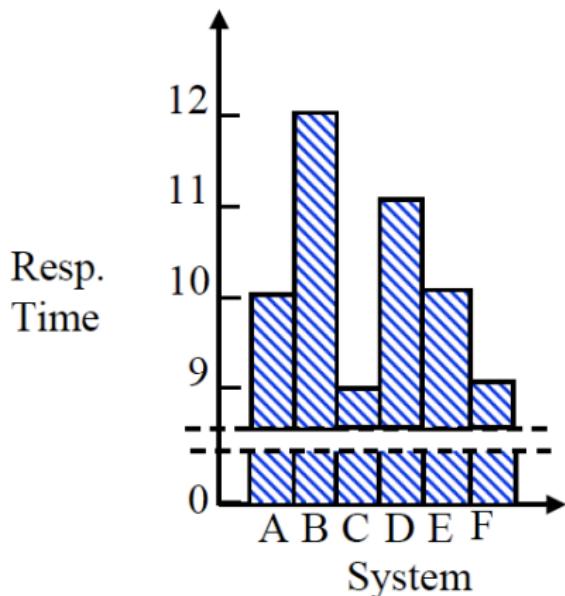
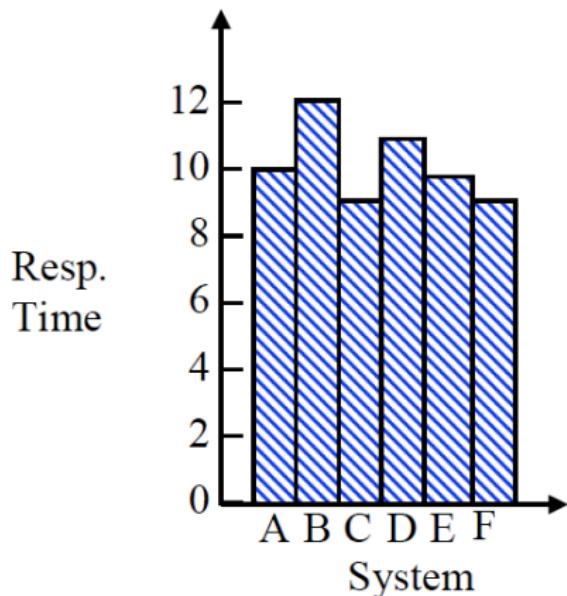
# COMMON MISTAKES

## Howto cheat ?



# COMMON MISTAKES

Howto cheat ?



# DATA

- ▶ The type of the graphic is adapted to the nature of data (curve, bars, pie, histogram, cloud...);
- ▶ Approximations/interpolation make sense ;
- ▶ Curves are defined by a sufficient number of points ;
- ▶ The building method of the curve is clear : interpolation (linear, polynomial, regression...);
- ▶ Confidence intervals are visualized (or given separately) ;
- ▶ Steps of histograms are adequate ;
- ▶ Histograms visualize probabilities (from 0 to 1).

The nature of the data implies the type of representation

# GRAPHICAL OBJECTS

- ▶ Graphical objects are readable on screen, on printed version (B/W), on video... ;
- ▶ Graphic range is standard, without too similar colors, without green (video) ;
- ▶ Graphical axis are well identified and labelled ;
- ▶ Scales and units are explicits ;
- ▶ Curves cross without ambiguity ;
- ▶ Grids help the reader.

Graphical objects provide the readability of the graphic

# ANNOTATIONS

- ▶ Axis are labelled by quantities ;
- ▶ Labels of the axis are clear, and self contained ;
- ▶ Units are indicated on the axis ;
- ▶ Axes are oriented from the left to the right and from the bottom to the top ;
- ▶ Origin is  $(0, 0)$ , if not it should be clearly justified ;
- ▶ No hole on the axes.
- ▶ For bar graphs/histograms order of bars is based on classical ordering (alphabetical, temporal, from the best to the worse) are better than a random order ;
- ▶ Each curve has a legend ;
- ▶ Each bar has a legend ;

Annotations put a semantic on graphics

# INFORMATION

- ▶ Curves are on the same scale ;
- ▶ The number of curves on a same graph is small (less than 6) ;
- ▶ Compare curves on a same graphic ;
- ▶ A curve cannot be removed without reducing the information ;
- ▶ The graphic gives a relevant information to the reader ;
- ▶ If the vertical axis shows averages, it should indicate error bars ;
- ▶ It is not possible to remove any object without modifying the readability of the graphic.

**Graphical information should answer some precise question**

# CONTEXT

- ▶ All the symbols are defined and referenced in the text ;
- ▶ The graphic produces more information than any other representation (choice of the variable) ;
- ▶ The graphic has a title ;
- ▶ The title is sufficiently self contained to partially understand the graphic ;
- ▶ The graphic is referenced in the text ;
- ▶ The text comment the figure.

A graphic should be a partial necessary information in a specific context

# SYNTHESIS

Keep always in mind : Who is the reader and why should he read the graphic ?

Hints for the design of a good graphical representation.

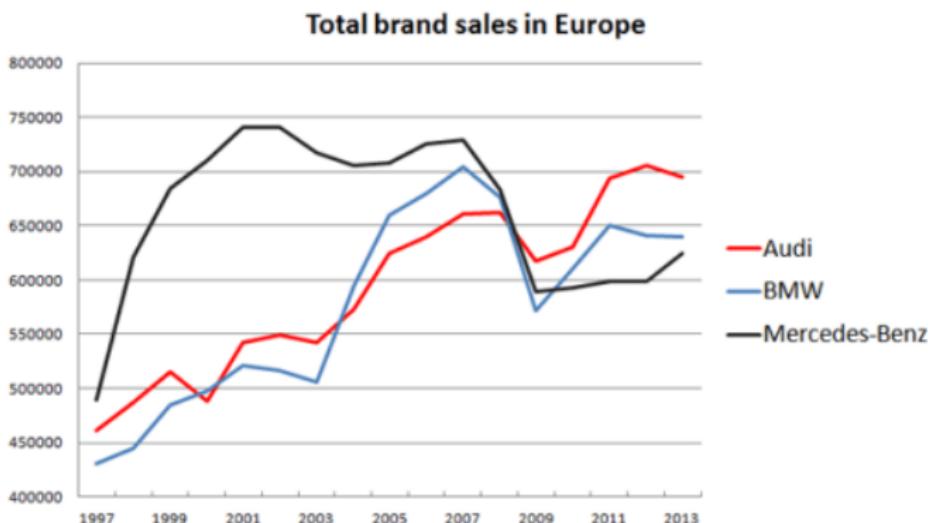
- ▶ Minimize efforts of the reader ;
- ▶ Maximize information ;
- ▶ Minimize *ink* ;
- ▶ Use traditional conventions
- ▶ Make several representations, before choosing the more adequate.
- ▶ Some classical errors
  - ▶ Too many graphical objects
  - ▶ Confusing scales, Cryptic notations
  - ▶ Non necessary informations,

## Principles

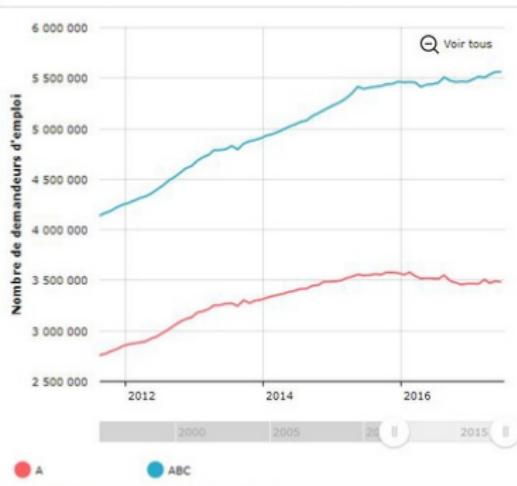
- ▶ **Occam's Razor** If two representations contain the same information, choose the simpler one.
- ▶ **Completion (Dijkstra)** When you cannot remove any simple object from the representation, then it is complete.
- ▶ **Common sense** Use an adapted sophistication level.

From Jean-Yves Le Boudec.

Last but not least : **The graphical representation should be elegant**



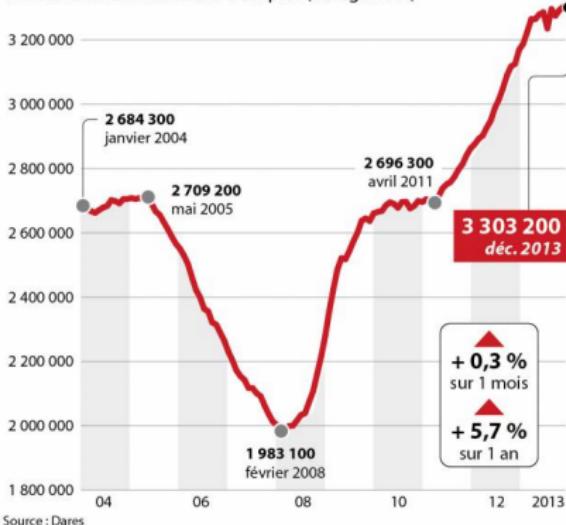
Demandeurs d'emploi inscrits en fin de mois à Pôle emploi, Catégorie A,  
ABC - France métropolitaine - Janvier 1996 à Juin 2017 

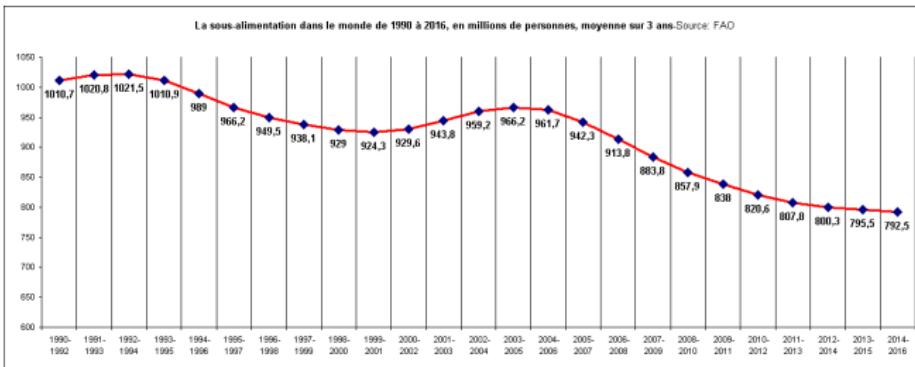


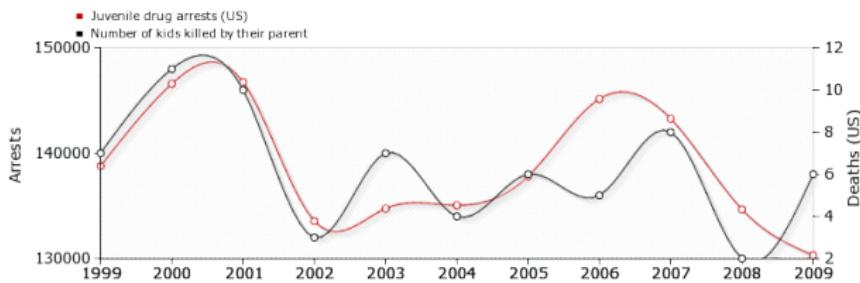
Source : Pôle emploi -Dares, STMT, Données CVS-CJO.

## Le chômage

Nombre de demandeurs d'emploi (catégorie A)



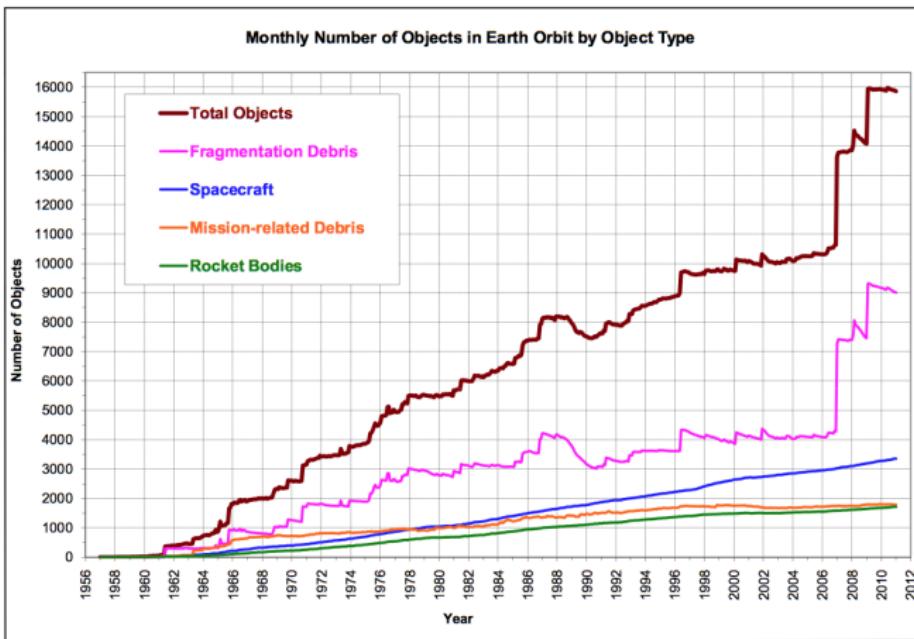






Investing.com





*Monthly Number of Cataloged Objects in Earth Orbit by Object Type:* This chart displays a summary of all objects in Earth orbit officially cataloged by the U.S. Space Surveillance Network. "Fragmentation debris" includes satellite breakup debris and anomalous event debris, while "mission-related debris" includes all objects dispensed, separated, or released as part of the planned mission.

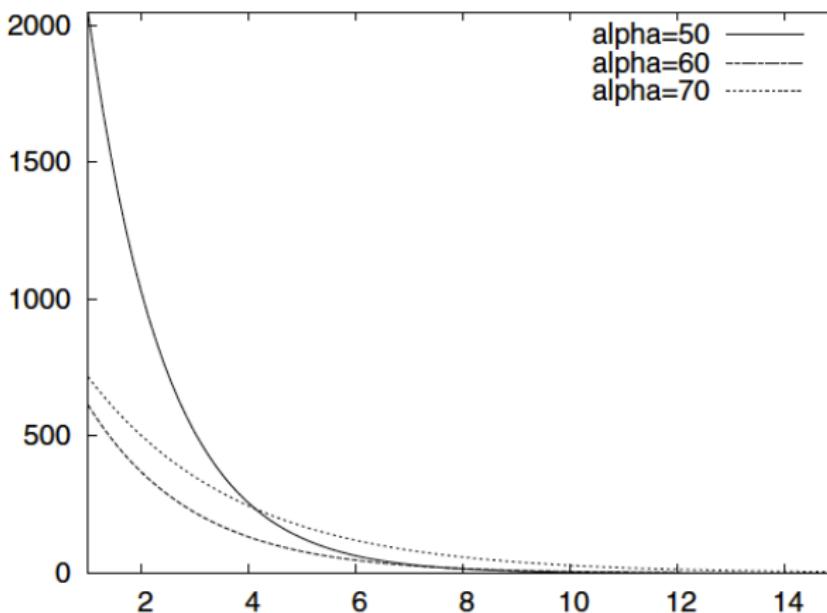
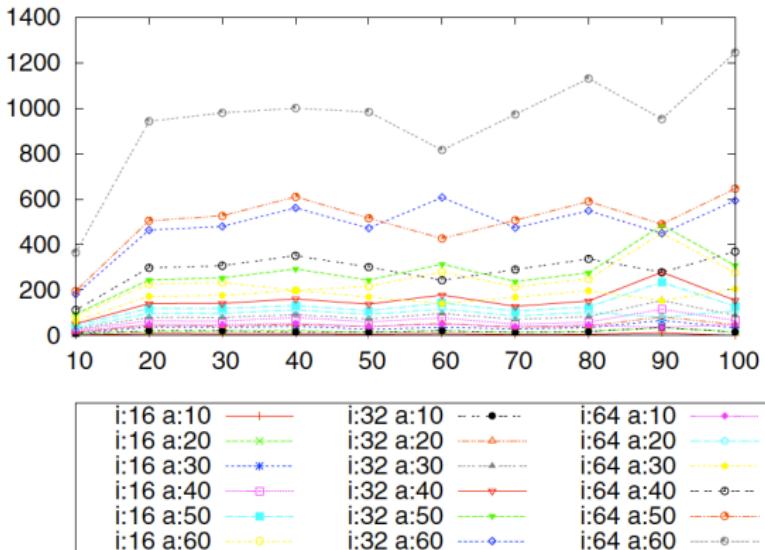


Figure 1. Temperature cooling scheme for different  $\alpha$  values

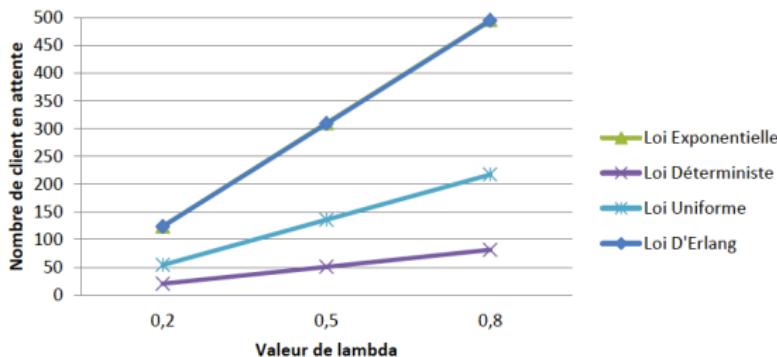


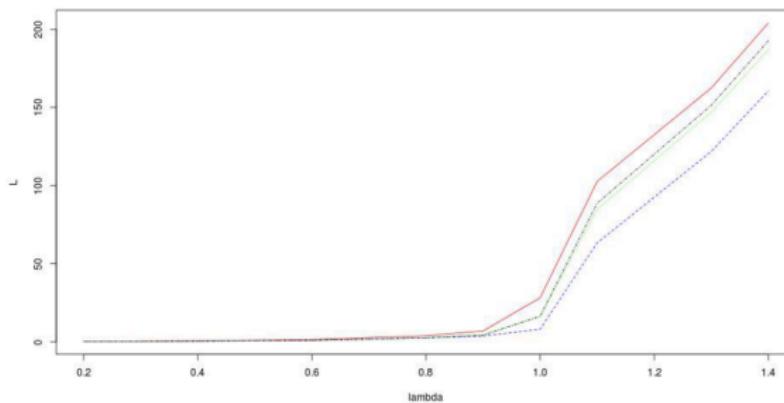
**Figure 2. SA – 512-node grid – pdf net 10 – i stands for iterations and a for  $\alpha$**

**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	$517.51 \pm 1.00E5$	$350.85 \pm 1.0E4$	$362.64 \pm 3.44E4$	$599.41 \pm 167523.40$

## Nombre moyen de clients dans la file en fonction de lambda

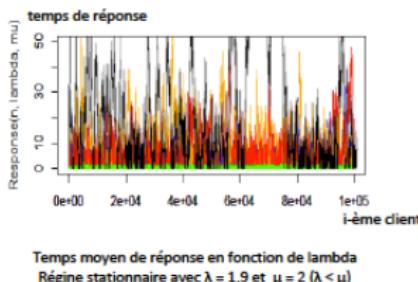




### Légende

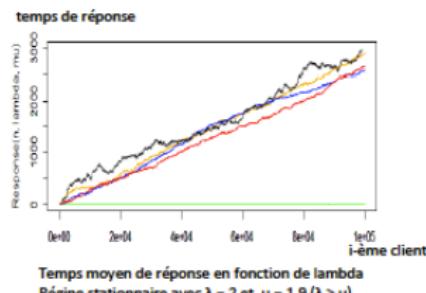
- Rouge : loi exponentielle
- Bleu : loi uniforme comprise entre 0 et 2
- Vert : loi uniforme entre 0,5 et 1,5
- Noir : loi d'Erlang.

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.

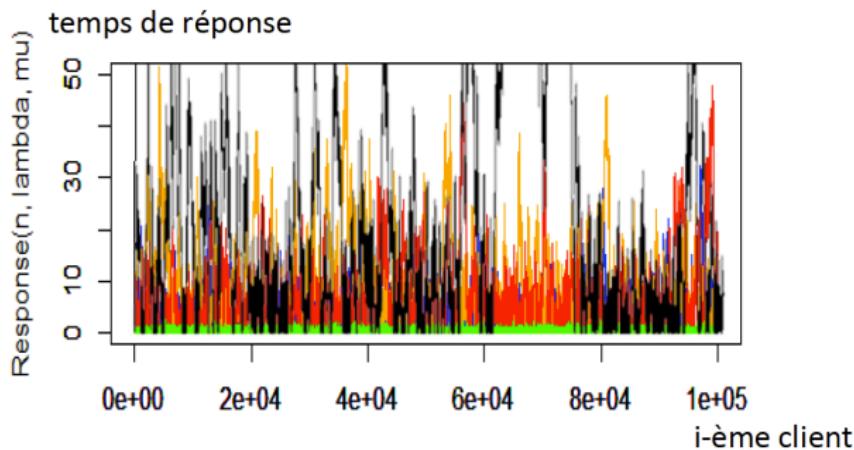
Nous lançons à présent une simulation pour chaque loi avec un taux d'arrivée moyen  $\lambda=2$  et un taux de service moyen  $\mu=1,9$ . C'est la saturation ( $\lambda > \mu$ ) :



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 régénération.



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

## Screen-shots

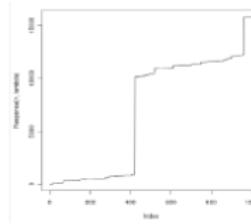


Figure 1 - Loi exponentielle

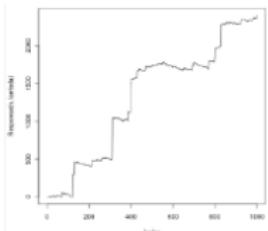


Figure 2 - loi uniforme sur l'intervalle  $[0;2]$

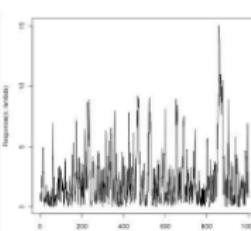


Figure 3 - loi uniforme sur l'intervalle  $[0.5;1.5]$

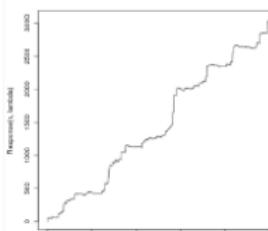


Figure 4 - Loi d'Erlang de paramètres  $(1,1)$

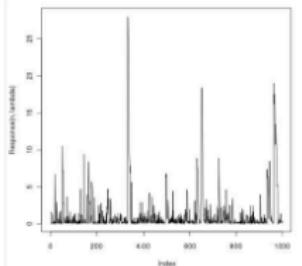


Figure 5 - Loi d'Erlang de paramètre  $(2,0.5)$

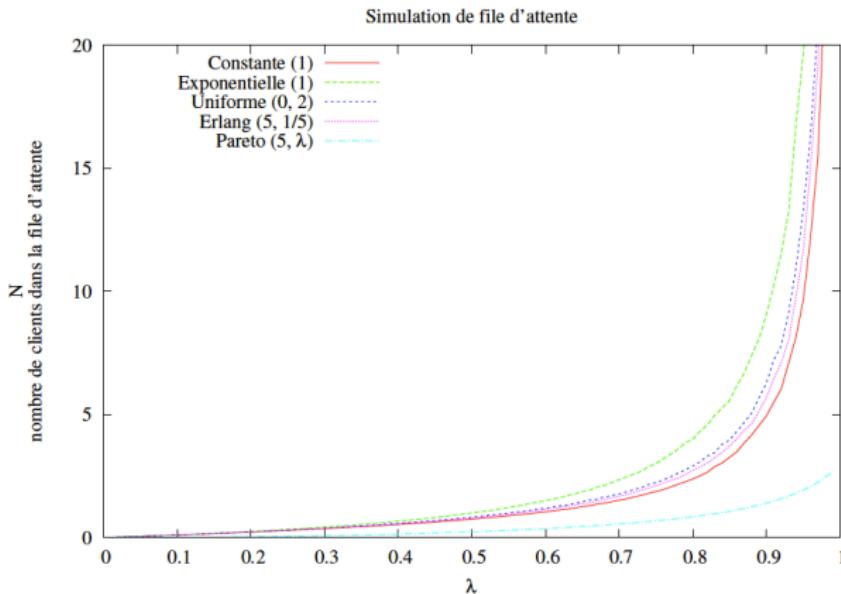


FIGURE 1 – Impact de la variabilité du temps de service sur le nombre moyen de clients dans la file d'attente

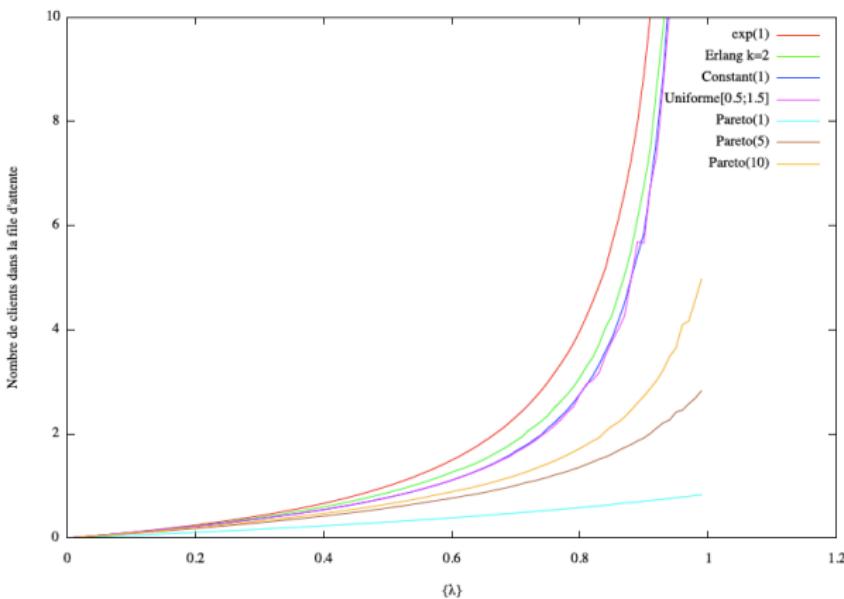


Figure : Evolution de la file d'attente en fonction du débit (arrivée et traitement)