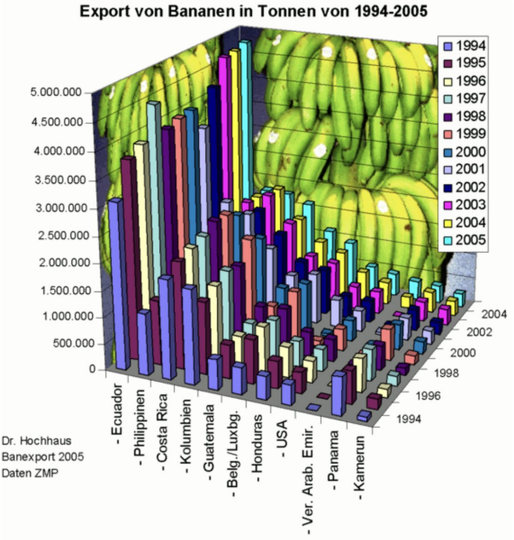
What a nice picture !

Data Visualization, an introduction

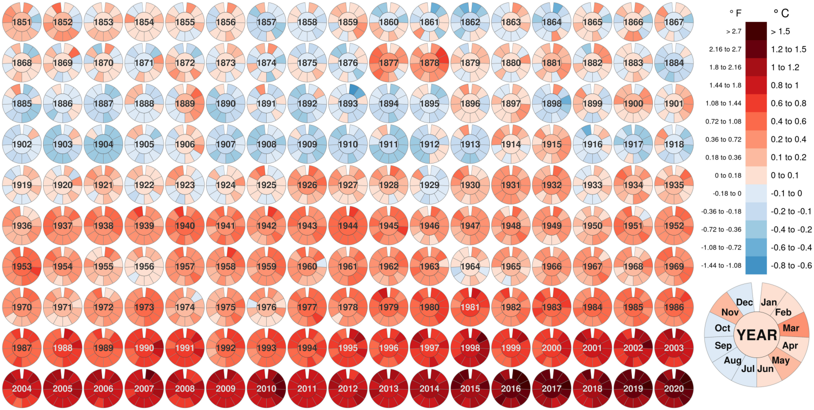
{Elise.Arnaud,Arnaud.Legrand,Jean-Marc.Vincent}@univ-grenoble-alpes.fr

Université de Grenoble-Alpes, UFR IM^{2}AG
M2R-MOSIG Scientific Methodology

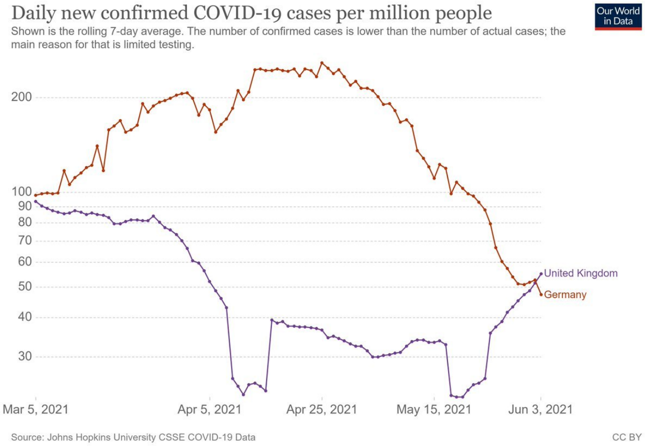
October 2022

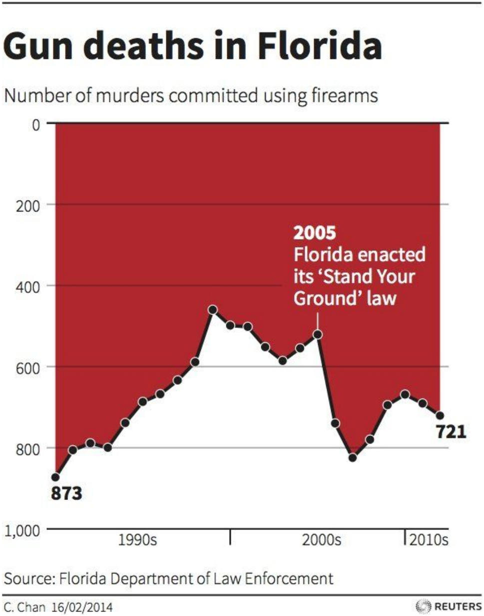

Export von Bananen in Tonnen von 1994-2005
5.000.000
4.500.000
4.000.000
3.500.000
3.000.000
2.500.000
2.000.000
1.500.000
1.000.000
500.000
0
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
- Ecuador
- Philippinen
- Costa Rica
- Kolumbien
- Guatemala
- Belg./Luxbg.
- Honduras
- USA
- Ver. Arab. Emir.
- Panama
- Kamerun
Dr. Hochhaus
Banexport 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
200
100
90
80
70
60
50
40
30
United Kingdom
Germany
Mar 5, 2021
Apr 5, 2021
Apr 25, 2021
May 15, 2021
Jun 3, 2021
Source: Johns Hopkins University CSSE COVID-19 Data
CC BY


Gun deaths in Florida
Number of murders committed using firearms
0
200
400
600
800
1,000
2005
Florida enacted
its 'Stand Your
Ground' law
873
721
1990s
2000s
2010s
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

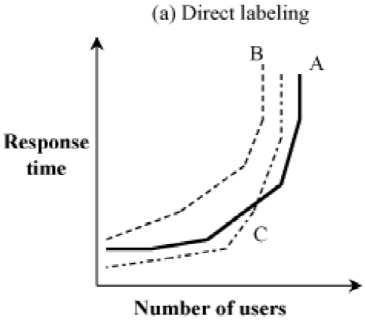
- **Increases deeply the quality of a paper**
- Show the scientific quality of your research
- Observation leads to open problems
- Pictures generates discussions

Mistakes

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

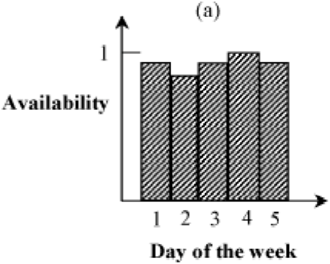
GUIDELINES FOR GOOD GRAPHICS (JAIN)

Minimum effort for the reader



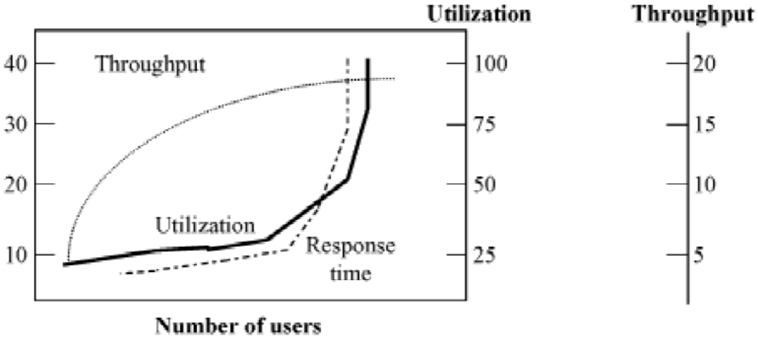
GUIDELINES FOR GOOD GRAPHICS (JAIN)

Minimize Ink



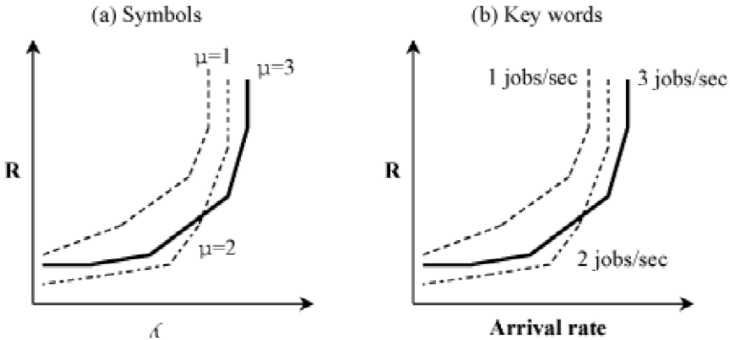
COMMON MISTAKES

Multiple scaling, Too much information



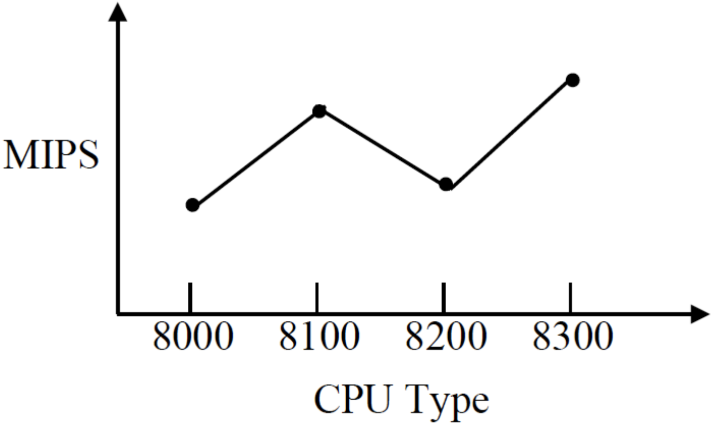
COMMON MISTAKES

Cryptic information



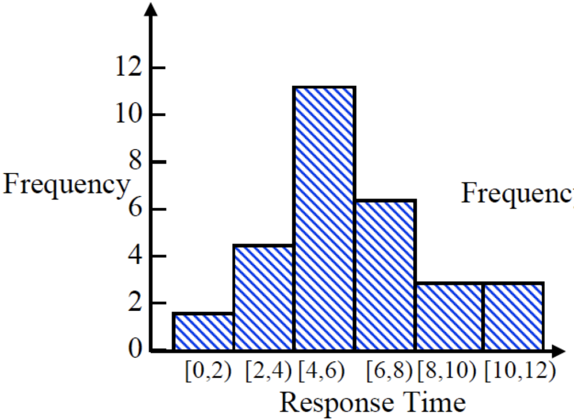
COMMON MISTAKES

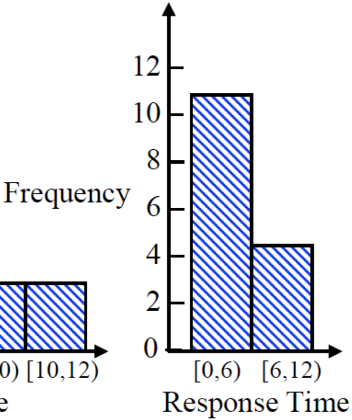
Non-relevant graphic objects



Common mistakes

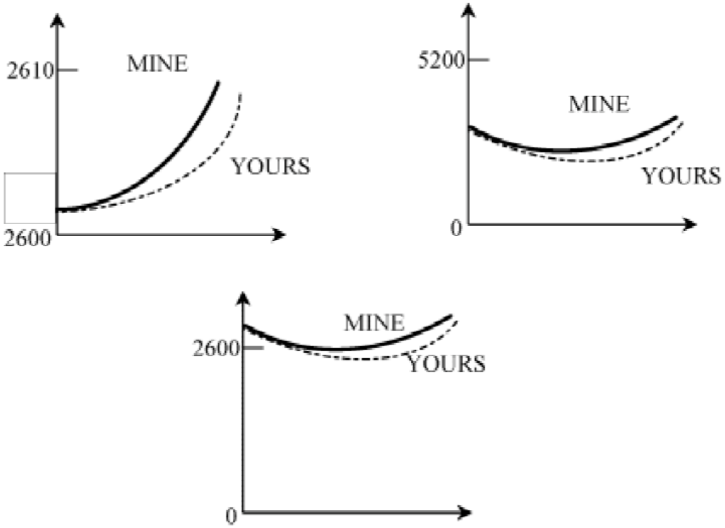
Non-relevant graphic objects





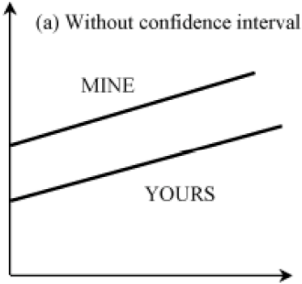
COMMON MISTAKES

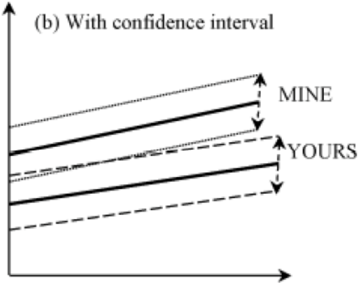
Howto cheat ?



COMMON MISTAKES

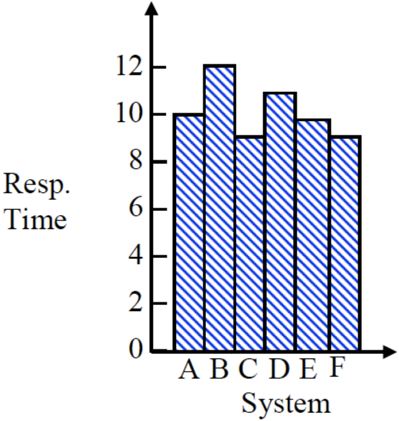
Howto cheat ?

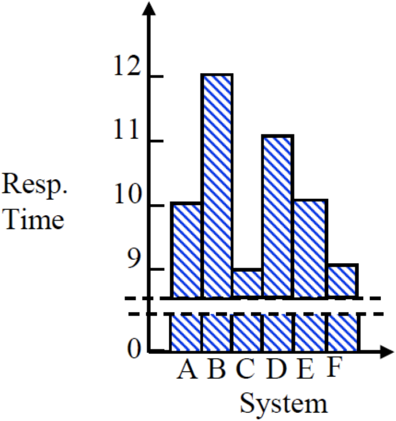




COMMON MISTAKES

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 indicates error bars ;
- It is not possible to remove any objet 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,

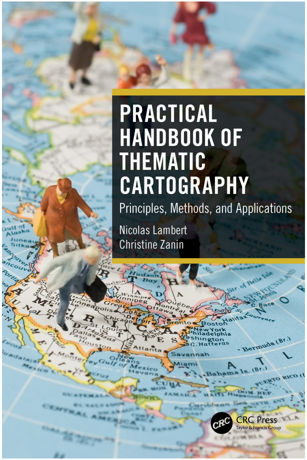
Principles

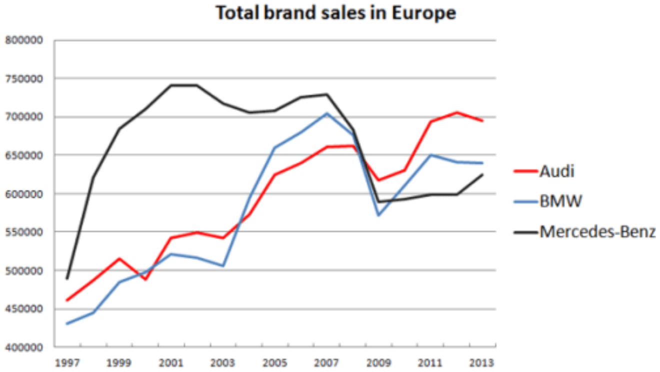
- **Occam's Razor** If two representations contain the same information, choose the simpler one.
- **Completion (Dijkstra)** When you cannot remove any simple object from the representation, then it is complete.
- **Common sense** Use an adapted sophistication level.

From Jean-Yves Le Boudec.

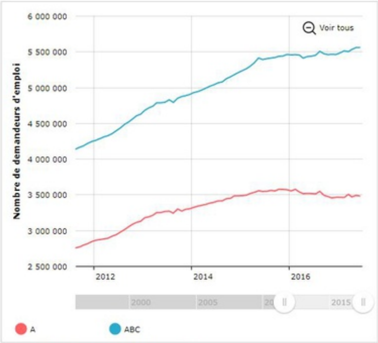
Last but not least : **The graphical representation should be elegant**

A Nice Reference

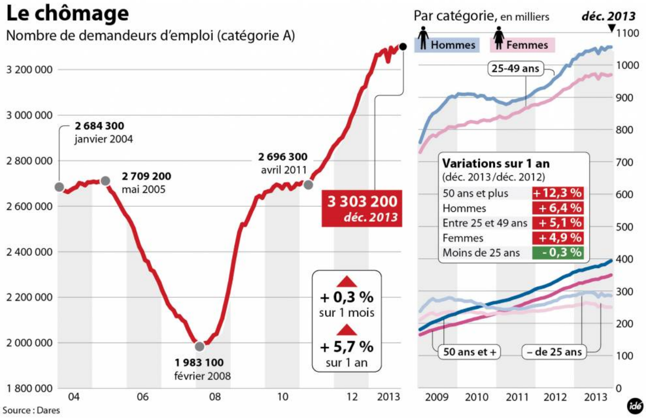


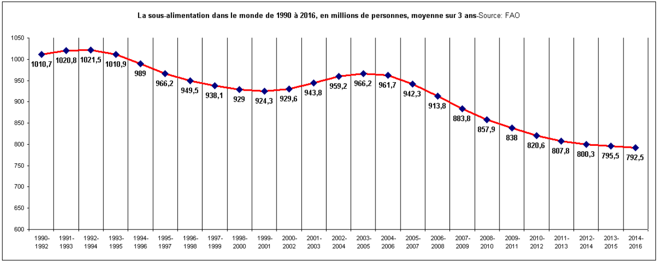

Total brand sales in Europe
800000
750000
700000
650000
600000
550000
500000
450000
400000
1997
1999
2001
2003
2005
2007
2009
2011
2013
Audi
BMW
Mercedes-Benz

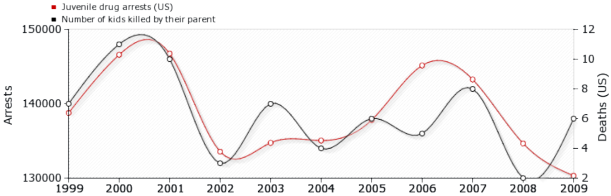
Demandeurs d'emploi inscrits en fin de mois à Pôle emploi, Catégorie A, ABC - France métropolitaine - Janvier 1996 à Juin 2017



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

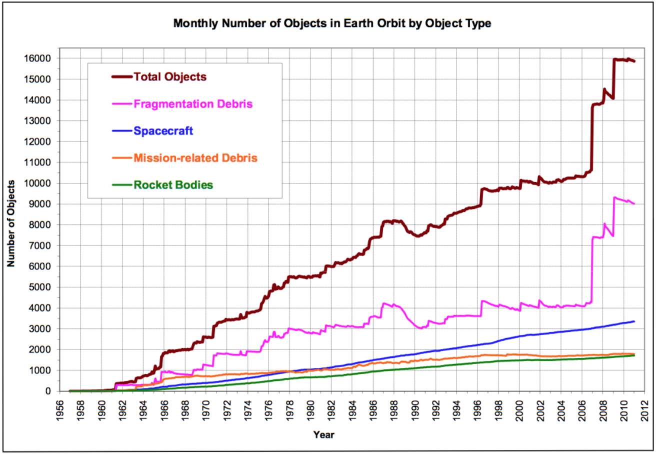

Le chômage
Nombre de demandeurs d'emploi (catégorie A)
3 200 000
3 000 000
2 800 000
2 600 000
2 400 000
2 200 000
2 000 000
1 800 000
2 684 300
janvier 2004
2 709 200
mai 2005
1 983 100
février 2008
2 696 300
avril 2011
3 303 200
déc. 2013
+ 0,3 %
sur 1 mois
+ 5,7 %
sur 1 an
04
06
08
10
12
2013
Source : Dares
Par catégorie, en milliers
déc. 2013
Hommes
Femmes
25-49 ans
Variations sur 1 an
(déc. 2013/déc. 2012)
50 ans et plus +12,3 %
Hommes +6,4 %
Entre 25 et 49 ans +5,1 %
Femmes +4,9 %
Moins de 25 ans -0,3 %
50 ans et +
– de 25 ans
1100
1000
900
800
700
600
500
400
300
200
100
0
2009
2010
2011
2012
2013


La sous-alimentation dans le monde de 1990 à 2016, en millions de personnes, moyenne sur 3 ans-Source: FAO
1050
1000
950
900
850
800
750
700
650
600
1010,7
1020,8
1021,5
1010,9
989
966,2
949,5
938,1
929
924,3
929,6
943,8
959,2
966,2
961,7
942,3
913,8
883,8
857,9
838
820,6
807,8
800,3
795,5
792,5
1990-1992
1991-1993
1992-1994
1993-1995
1994-1996
1995-1997
1996-1998
1997-1999
1998-2000
1999-2001
2000-2002
2001-2003
2002-2004
2003-2005
2004-2006
2005-2007
2006-2008
2007-2009
2008-2010
2009-2011
2010-2012
2011-2013
2012-2014
2013-2015
2014-2016


Juvenile drug arrests (US)
Number of kids killed by their parent
Arrests
Deaths (US)
150000
140000
130000
12
10
8
6
4
2
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009


XRP/EUR, 15, BitStamp
O 0.98500 H 0.99419 B 0.97319 C 0.97380
Market Open
1.60000
1.50000
1.40000
1.30000
1.20000
1.10000
1.00000
0.97380
0.90000
0.80000
0.70000
Investing.com
12:00
16
12:00
17


ÉVOLUTION DU NOMBRE DE CATASTROPHES NATURELLES DANS LE MONDE
SOURCE : MUNICH RE
800
600
400
200
1980
1990
2000
2010
2017



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.

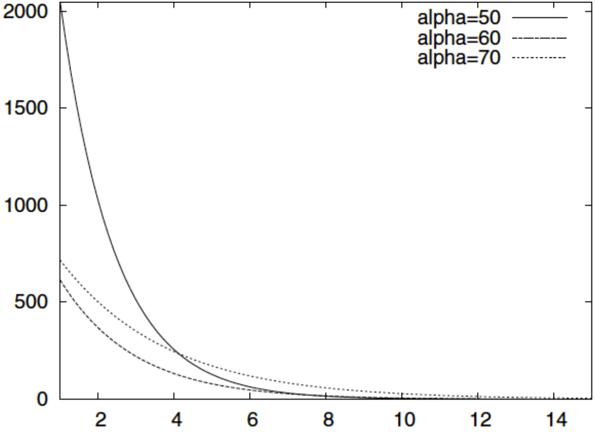


Figure 1. Temperature cooling scheme for different α values

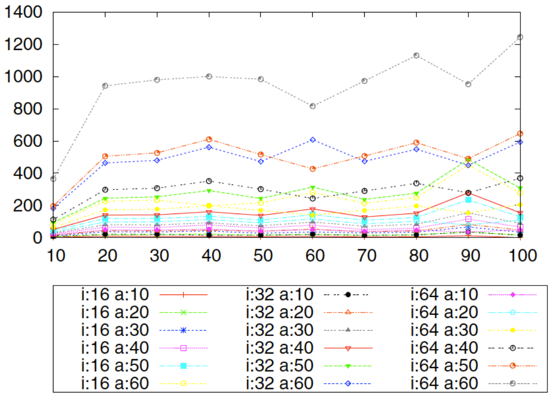
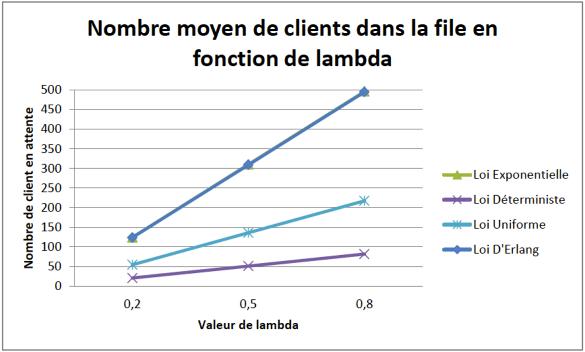
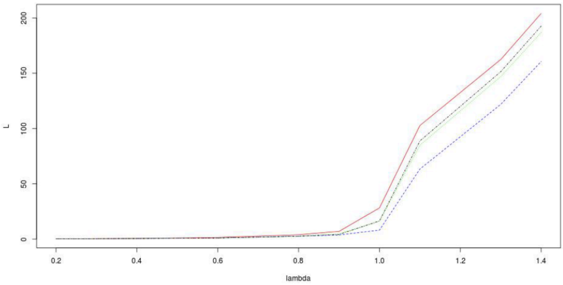


Figure 2. SA – 512-node grid – pdf net 10 – i stands for iterations and a for α

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

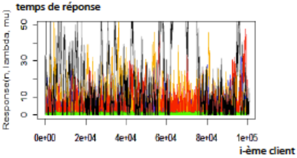
#Apps	Lowest	SA	Route	Round Robin
	MRT–Stdev	MRT–Stdev	MRT–Stdev	MRT–Stdev
10	$113.75 \pm 8.65E3$	$62.89 \pm 1.4E3$	$110.39 \pm 1.01E4$	129.62 ± 10890.67
20	$212.71 \pm 1.66E4$	$68.40 \pm 1.2E3$	$205.04 \pm 1.66E4$	226.10 ± 19521.60
30	$324.89 \pm 7.33E4$	$71.50 \pm 2.5E3$	$213.69 \pm 2.41E4$	262.34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


Nombre moyen de clients dans la file en fonction de lambda
Nombre de client en attente
500
450
400
350
300
250
200
150
100
50
0
0,2
0,5
0,8
Valeur de lambda
Loi Exponentielle
Loi Déterministe
Loi Uniforme
Loi D'Erlang



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

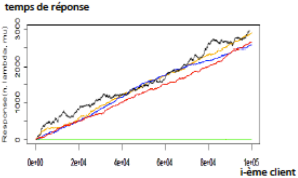
Nous lançons à présent une simulation pour chaque loi avec un taux d'arrivée moyen λ=1,9 et un taux de service moyen μ=2. Le régime est stable : (λ < μ) :



Temps moyen de réponse en fonction de lambda
Régine stationnaire avec λ = 1,9 et μ = 2 (λ < μ)

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 λ=2 et un taux de service moyen μ=1,9. C'est la saturation (λ > μ). :



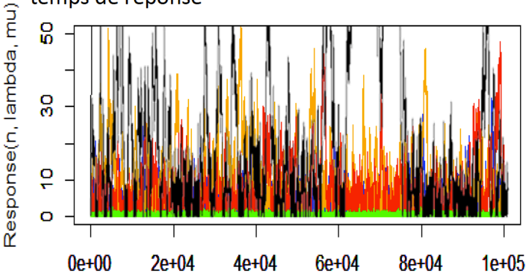
Temps moyen de réponse en fonction de lambda
Régine stationnaire avec λ = 2 et μ = 1,9 (λ > μ)

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.

temps de réponse



i-ème client

Temps moyen de réponse en fonction de lambda
Régine stationnaire avec λ = 1,9 et μ = 2 (λ < μ)

Screen-shots

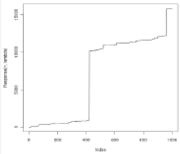


Figure 1 - Loi exponentielle

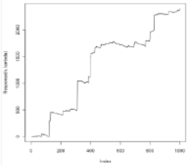


Figure 2 - loi uniforme sur l'intervalle [0;2]

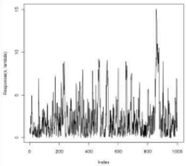


Figure 3 - loi uniforme sur l'intervalle [0.5;1.5]

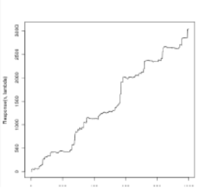


Figure 4 - Loi d'Erlang de paramètres (1,1)

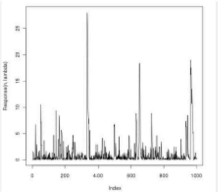


Figure 5 - Loi d'Erlang de paramètre (2,0.5)

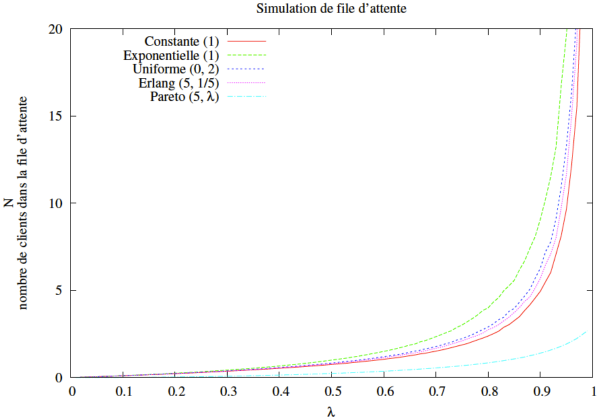


FIGURE 1 – Impact de la variabilité du temps de service sur le nombre moyen de clients dans la file d'attente

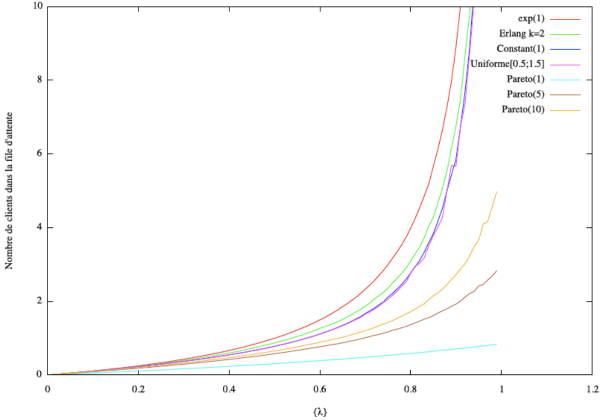


Figure : Evolution de la file d'attente en fonction du débit (arrivée et traitement)