



PRÉSENTER SES DONNÉES GRAPHIQUEMENT

Marc Girondot

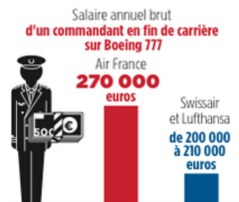
1

RÈGLES DE BASE

- Un graphique doit représenter fidèlement les données et ne pas chercher à masquer ou amplifier des informations
- Le type de graphique doit être approprié aux types de données
- Un graphique doit avoir des axes légendés avec des unités et un titre et une échelle graduée logique
- Un graphique ne doit pas utiliser de fioritures qui pourraient être sujettes à interprétation
- Si des barres d'erreur sont indiquées, il doit être clairement signifié ce qu'elles représentent

2

UN GRAPHIQUE DOIT ÊTRE HONNÊTE



Salaire annuel brut d'un commandant en fin de carrière sur Boeing 777

Air France **270 000 euros**

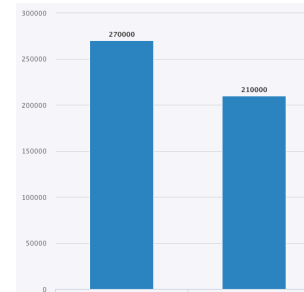
Swissair et Lufthansa de 200 000 à 210 000 euros

Le Point

- Ce graphique est-il honnête ?

3

UN GRAPHIQUE DOIT ÊTRE HONNÊTE



- On peut amplifier ou masquer une tendance en changeant l'échelle ou l'origine du graphique

4

HONNÊTETÉ

- Le degré d'honnêteté dépendra d'une information externe qui n'est pas incluse dans les données elle-même.
- Une telle différence peut être très importante ou au contraire purement secondaire.
 - Une façon de répondre à la question est de tester si la pente est significativement différente de 0;
 - Mais attention encore, une significativité statistique peut ne pas représenter une significativité biologique, voir par exemple Yoccoz, N.G., 1991. Use, overuse, and misuse of significance tests in evolutionary biology and ecology. Bulletin of the Ecological Society of America 72, 106-111.

5

LE TYPE DE GRAPHIQUE

- Données quantitatives et qualitatives
 - Quantitatives: 1; 3; 9; 15
 - Qualitatives: 1 2 3 4 ou A B C D
- Données quantitatives continues, discrètes ou discrétisées
 - Continue: 1; 1,1; 1,01; 1,0003
 - Discrète: 1; 2; 3; 4 ou 10; 20; 30
 - Discrétisées: [0; 1[[1; 2[

6

Deux données quantitatives

- Diagramme bivarié de type XY

7

Deux données quantitatives

- Relier les points ou non; si les données disponibles sont obtenues graduellement selon l'axe des X, on peut les relier (ex. temps, ou distance par rapport à une référence).

8

Deux données quantitatives

- Relier les points ou non; si les données disponibles sont obtenues graduellement selon l'axe des X, on peut les relier (ex. temps, ou distance par rapport à une référence).

9

Deux données quantitatives

- Relier les points ou non; si les données disponibles sont obtenues graduellement selon l'axe des X, on peut les relier (ex. temps, ou distance par rapport à une référence).

10

Points reliés ou non

- La différence n'est pas seulement esthétique. Il faut se demander si relier les points a un sens.
 - Si on relie les points par une droite, le modèle linéaire peut-il être justifié? Si c'est une courbe, prendre garde car les hypothèses sont aussi nombreuses.

11

UNE VARIABLE CONTINUE

- Si les données sont constituées d'une variable quantitative continue, on peut la représenter de différentes façons.

- Cette façon de présenter les données n'a aucun intérêt. Aucune information ne peut-être tirée de ce graphique.

12

UNE VARIABLE CONTINUE

- Si les données sont constituées d'une variable quantitatives continues, on peut trier les données selon la valeur des ordonnées.

13

UNE VARIABLE CONTINUE

- On peut en tirer différentes statistiques telles que médiane, quartiles ou déciles.

14

UNE VARIABLE CONTINUE

- On peut en tirer différentes statistiques telles que médiane, quartiles ou déciles.

15

UNE VARIABLE CONTINUE

- On peut en tirer différentes statistiques telles que médiane, quartiles ou déciles.

16

UNE VARIABLE CONTINUE

- On peut aussi dénombrer le nombre d'individus se trouvant dans un intervalle. On choisira préférentiellement des intervalles de même taille.
- Ex. Nombre entre 0 inclus et 10 exclus, $[0; 10[$
- Nombre entre 10 inclus et 20 exclus, $[10; 20[$
- Etc.

$[0; 10[$	=0
$[10; 20[$	=2
$[20; 30[$	=4
$[30; 40[$	=2
$[40; 50[$	=7
$[50; 60[$	=3
$[60; 70[$	=0
$[70; 80[$	=3
$[80; 90[$	=2
$[90; 100[$	=2
$[100; 110[$	=1

17

UNE VARIABLE CONTINUE

- Si les intervalles ne sont pas de même taille, il faut diviser la valeur de la catégorie pour se ramener à une proportionnalité de surface.

$[0; 10[$	=0
$[10; 20[$	=2
$[20; 30[$	=4
$[30; 40[$	=2
$[40; 50[$	=7
$[50; 60[$	=3
$[60; 70[$	=0
$[70; 80[$	=3
$[80; 90[$	=2
$[90; 100[$	=2
$[100; 110[$	=1

Faux

18

UNE VARIABLE CONTINUE

- Si les intervalles ne sont pas de même taille, il faut diviser la valeur de la catégorie pour se ramener à une proportionnalité de surface.

[0;10[=0	[0;10[0
[10;20[=2	[10;20[2
[20;30[=4	[20;30[4
[30;40[=2	[30;40[2
[40;50[=7	[40;50[7
[50;60[=3	[50;60[3
[60;70[=0	[60;70[0
[70;80[=3	[70;80[3
[80;90[=2	[80;90[2
[90;100[=2	[90;100[2
[100;110[=1	[100;110[1

8/5=0,6

19

EXEMPLE À NE PAS FAIRE

en milliards de dollars

Le Monde

Year	2002	2003	2004	2005	2006	2007	2008	2009
Value	0.4	0.1	0.4	1.8	3.1	4.2	4.2	1.4

20

UNE DONNÉE QUALITATIVE ET UNE QUANTITATIVE

- La donnée qualitative est mise sur l'axe X. Est-elle ordonnée ou non ? Si oui, faire attention à l'ordre des catégories.
- La donnée quantitative est-elle un dénombrement (diagramme en bâtons ou en points) ou une valeur (diagramme en points) ?

Usine	1	2	3	4	5	6	7	8	9	10	11	12
Productions	180	220	160	210	200	250	200	230	170	120	220	210

21

UNE DONNÉE QUALITATIVE ET UNE QUANTITATIVE

- La donnée qualitative est mis sur l'axe X. Est-elle ordonnée ou non ? Si oui, faire attention à l'ordre des catégories.
- La donnée quantitative est-elle un dénombrement (diagramme en bâtons ou en points) ou une valeur (diagramme en points) ?

Numéro de Usine	1	2	3	4	5	6	7	8	9	10	11	12
Valeur	100	120	110	130	120	140	130	110	100	120	130	140

22

EVITER LES FIORITURES QUI N'ONT PAS DE SIGNIFICATION

Comment interpréter le dégradé; comment interpréter la profondeur ?

Usine	1	2	3	4	5	6	7	8	9	10	11	12
Productions	180	220	160	210	200	250	200	230	170	120	220	210

23

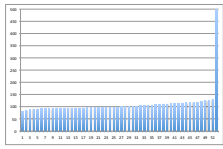
INDIQUER L'ERREUR SUR DES MESURES

- Si un point représente une collection de mesures, il est important de montrer la dispersion autour de cette valeur.
- Il existe différentes façon de mesurer la dispersion des valeurs.
 - Maximum, minimum
 - Multiple de l'écart-type
 - Multiple de l'erreur standard
 - Quartile, Décile, Centile

24

MAXIMUM, MINIMUM

- Soit un jeu de données:
 - La moyenne arithmétique = 109,27
 - Le maximum est 502,09
 - Le minimum est 78,20



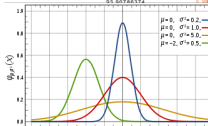
- Les maximums et minimums sont très sensibles aux valeurs extrêmes qui peuvent être aberrantes.

90,3022875	104,2087824
114,7314327	106,9744451
106,9744451	90,3022875
105,5617497	92,5674796
92,5674796	109,1282714
109,1282714	92,5674796
109,2009865	121,8742706
121,8742706	94,25174591
94,25174591	92,5170973
92,5170973	92,98157116
92,98157116	128,87118
128,87118	114,9991892
114,9991892	110,2004402
110,2004402	96,3972827
96,3972827	90,2084614
90,2084614	97,05399126
97,05399126	100,2009043
100,2009043	99,65821255
99,65821255	116,6020201
116,6020201	89,3388027
89,3388027	89,4432898
89,4432898	111,3529959
111,3529959	98,4432898
98,4432898	116,0476233
116,0476233	124,2512524
124,2512524	101,4738267
101,4738267	105,25861
105,25861	88,9438077
88,9438077	78,20880103
78,20880103	82,3218279
82,3218279	90,3178292
90,3178292	99,21191704
99,21191704	110,2044455
110,2044455	93,9078974
93,9078974	104,2087824
104,2087824	80,1073365
80,1073365	105,2477515
105,2477515	98,0884244
98,0884244	98,13814246
98,13814246	101,4242228
101,4242228	94,5042286
94,5042286	99,9421019
99,9421019	94,11716268
94,11716268	113,5505887
113,5505887	82,2008828
82,2008828	87,2708421
87,2708421	94,2007899

25

ECART-TYPE, STANDARD DEVIATION

- La moyenne arithmétique = 109,27
- L'écart-type = SD = 56,6
 - Racine carrée de la moyenne des carrés des écarts à la moyenne
- L'écart-type est une mesure de la dispersion des points.
- Cette statistique reste sensible aux valeurs extrêmes
 - ex., sans la valeur de 502,09, il n'est plus que de 11,36.
- Si la distribution est Gaussienne, 95% des points sont situés entre moyenne \pm 2.SD
 - Cela reste à peu près vrai si la distribution est unimodale et symétrique.



90,3022875	104,2087824
114,7314327	106,9744451
106,9744451	90,3022875
105,5617497	92,5674796
92,5674796	109,1282714
109,1282714	92,5674796
109,2009865	121,8742706
121,8742706	94,25174591
94,25174591	92,5170973
92,5170973	92,98157116
92,98157116	128,87118
128,87118	114,9991892
114,9991892	110,2004402
110,2004402	96,3972827
96,3972827	90,2084614
90,2084614	97,05399126
97,05399126	100,2009043
100,2009043	99,65821255
99,65821255	116,6020201
116,6020201	89,3388027
89,3388027	89,4432898
89,4432898	111,3529959
111,3529959	98,4432898
98,4432898	116,0476233
116,0476233	124,2512524
124,2512524	101,4738267
101,4738267	105,25861
105,25861	88,9438077
88,9438077	78,20880103
78,20880103	82,3218279
82,3218279	90,3178292
90,3178292	99,21191704
99,21191704	110,2044455
110,2044455	93,9078974
93,9078974	104,2087824
104,2087824	80,1073365
80,1073365	105,2477515
105,2477515	98,0884244
98,0884244	98,13814246
98,13814246	101,4242228
101,4242228	94,5042286
94,5042286	99,9421019
99,9421019	94,11716268
94,11716268	113,5505887
113,5505887	82,2008828
82,2008828	87,2708421
87,2708421	94,2007899

26

ERREUR STANDARD, STANDARD ERROR

- La moyenne arithmétique = 109,27
- L'erreur standard = 1,08
 - Racine carrée de la moyenne des carrés des écarts à la moyenne divisé par N
- L'erreur standard est une mesure de la précision avec laquelle est connue la moyenne
- Cette statistique reste sensible aux valeurs extrêmes
- Il y a 95% de chance que la vraie moyenne soit située entre moyenne \pm 2.SE (conclusion assez robuste à la forme de la distribution)

90,3022875	104,2087824
114,7314327	106,9744451
106,9744451	90,3022875
105,5617497	92,5674796
92,5674796	109,1282714
109,1282714	92,5674796
109,2009865	121,8742706
121,8742706	94,25174591
94,25174591	92,5170973
92,5170973	92,98157116
92,98157116	128,87118
128,87118	114,9991892
114,9991892	110,2004402
110,2004402	96,3972827
96,3972827	90,2084614
90,2084614	97,05399126
97,05399126	100,2009043
100,2009043	99,65821255
99,65821255	116,6020201
116,6020201	89,3388027
89,3388027	89,4432898
89,4432898	111,3529959
111,3529959	98,4432898
98,4432898	116,0476233
116,0476233	124,2512524
124,2512524	101,4738267
101,4738267	105,25861
105,25861	88,9438077
88,9438077	78,20880103
78,20880103	82,3218279
82,3218279	90,3178292
90,3178292	99,21191704
99,21191704	110,2044455
110,2044455	93,9078974
93,9078974	104,2087824
104,2087824	80,1073365
80,1073365	105,2477515
105,2477515	98,0884244
98,0884244	98,13814246
98,13814246	101,4242228
101,4242228	94,5042286
94,5042286	99,9421019
99,9421019	94,11716268
94,11716268	113,5505887
113,5505887	82,2008828
82,2008828	87,2708421
87,2708421	94,2007899

27

QUARTILE, DÉCILE ET CENTILE

- 1^{er} quartile donne la valeur pour laquelle 25% des individus sont dessous
- 2^{ème} quartile donne la valeur pour laquelle 50% des individus sont dessous
- 3^{ème} quartile donne la valeur pour laquelle 75% des individus sont dessous

28

EXEMPLE

- 1^{er} quartile = 93,65
- 3^{ème} quartile = 110,20
- Mesure très peu sensible aux valeurs extrêmes
 - Si bien sûr il y a moins de 25% de valeurs extrêmes faibles ou fortes !
- Peut-être utilisé même si la distribution est très asymétrique mais alors les barres d'erreur seront asymétriques.

78,20880103	1,82%
82,2008828	3,87%
87,2708421	5,77%
88,31488463	7,89%
88,8288027	9,82%
89,4432898	11,24%
90,2084614	12,45%
90,27238487	13,38%
90,8438875	17,31%
92,3218279	19,27%
92,3218279	21,15%
92,7288771	23,05%
93,8587478	25,00%
94,087824	26,82%
94,98157116	28,85%
96,1170888	30,77%
94,25174591	32,69%
94,2007899	34,62%
94,5842288	36,54%
95,2178292	38,46%
96,5885914	40,38%
96,2879237	42,31%
97,05399126	44,23%
97,9421019	46,15%
98,4432898	48,08%
98,2845494	50,00%
98,2845494	51,92%
98,2845494	53,84%
98,2121224	55,72%
101,4738267	59,82%
104,2087824	61,84%
105,25861	63,46%
105,2876487	65,38%
106,7477515	67,31%
106,784451	69,23%
108,1282714	71,15%
109,2009865	73,08%
110,2044455	75,00%
111,3529959	76,92%
113,5505887	78,85%
113,5505887	80,77%
114,7314327	82,69%
116,0885182	84,62%
116,0476233	86,54%
116,0885182	88,46%
121,8742706	92,31%
124,2512524	94,23%
124,7610574	96,15%
126,87118	98,08%
502,090043	100,00%

29