



PRÉSENTER SES DONNÉES GRAPHIQUEMENT

Marc Girondot

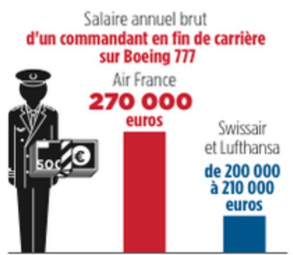
1

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

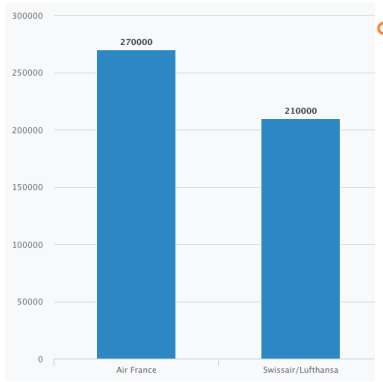
Swissair
et Lufthansa
de 200 000
à 210 000
euros

Le Point

- Ce graphique est-il honnête ?

3

UN GRAPHIQUE DOIT ÊTRE HONNÊTE



Compagnie	Salaires (euros)
Air France	270 000
Swissair/Lufthansa	210 000

- 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

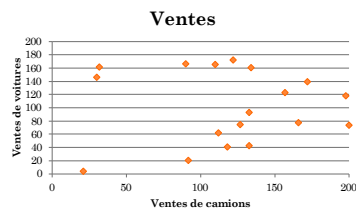
LE TYPE DE GRAPHIQUE

- 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

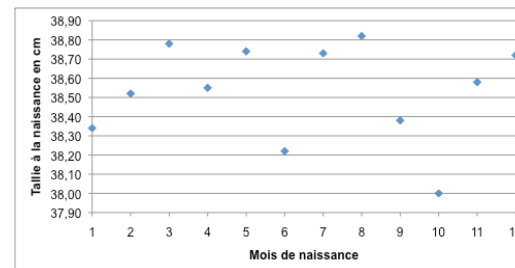
- Diagramme bivarié de type XY



7

Deux données quantitatives

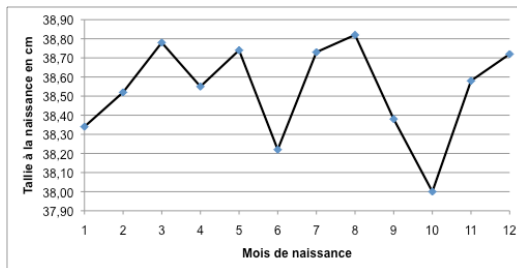
- Relier les points ou non; si les données disponibles sont obtenues graduellement selon l'axe des X, on peut les relier (ex. temps, ou distance par rapport à une référence).



8

Deux données quantitatives

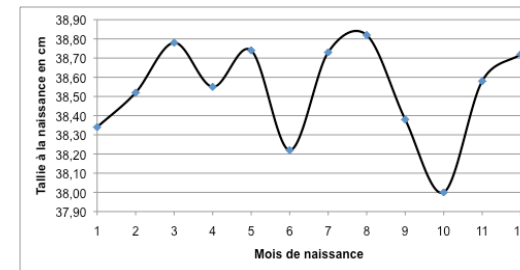
- Relier les points ou non; si les données disponibles sont obtenues graduellement selon l'axe des X, on peut les relier (ex. temps, ou distance par rapport à une référence).



9

Deux données quantitatives

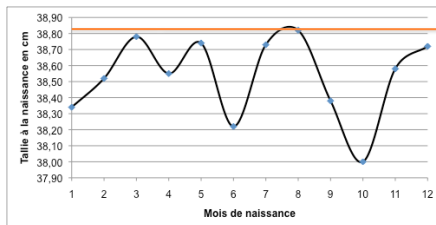
- Relier les points ou non; si les données disponibles sont obtenues graduellement selon l'axe des X, on peut les relier (ex. temps, ou distance par rapport à une référence).



10

Points reliés ou non

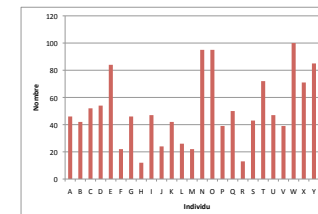
- La différence n'est pas seulement esthétique. Il faut se demander si relier les points a un sens.
 - Si on relie les points par une droite, le modèle linéaire peut-il être justifié ? Si c'est une courbe, prendre garde car les hypothèses sont aussi nombreuses.



11

UNE VARIABLE CONTINUE

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

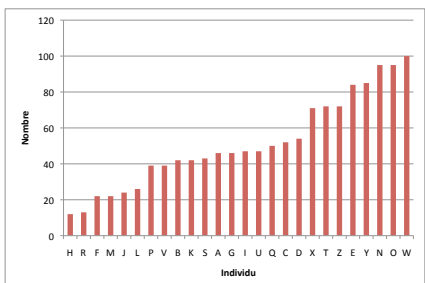


- Cette façon de présenter les données n'a aucun intérêt. Aucune information ne peut-être tirée de ce graphique.

12

UNE VARIABLE CONTINUE

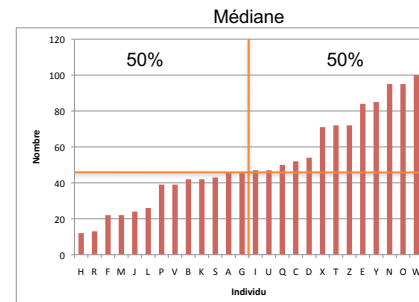
- Si les données sont constituées d'une variable quantitatives continues, on peut trier les données selon la valeur des ordonnées.



13

UNE VARIABLE CONTINUE

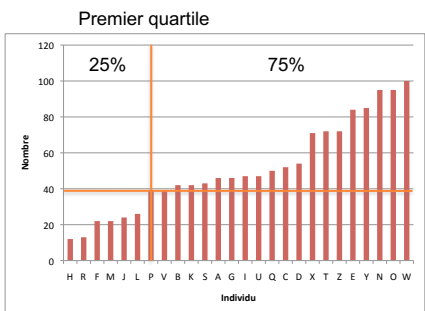
- On peut en tirer différentes statistiques telles que médiane, quartiles ou déciles.



14

UNE VARIABLE CONTINUE

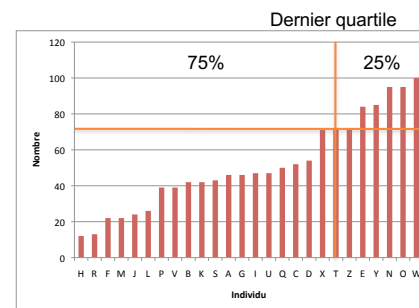
- On peut en tirer différentes statistiques telles que médiane, quartiles ou déciles.



15

UNE VARIABLE CONTINUE

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

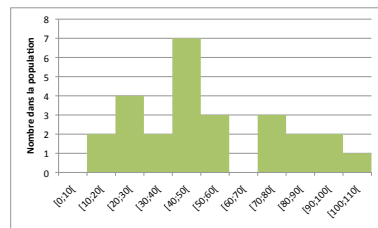


16

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



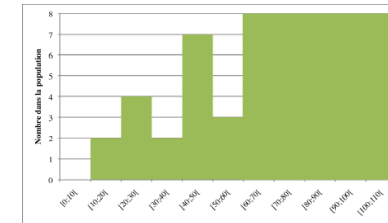
17

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

$[0;10[$	0
$[10;20[$	2
$[20;30[$	4
$[30;40[$	2
$[40;50[$	7
$[50;60[$	3
$[60;110[$	8



Faux

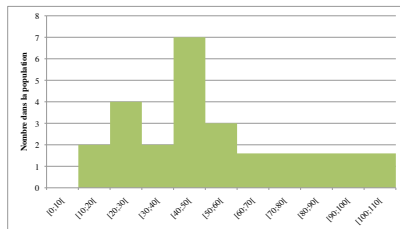
18

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
$[10;20[$	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

8/5=1,6

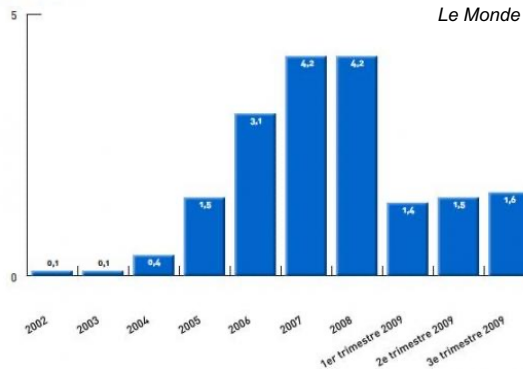


19

EXEMPLE À NE PAS FAIRE

en milliards de dollars

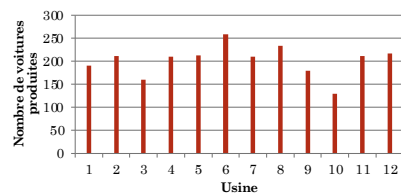
Le Monde



20

UNE DONNÉE QUALITATIVE ET UNE QUANTITATIVE

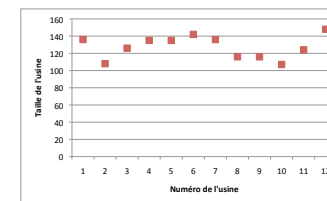
- La donnée qualitative est mise sur l'axe X. Est-elle ordonnée ou non ? Si oui, faire attention à l'ordre des catégories.
- La donnée quantitative est-elle un dénombrement (diagramme en bâtons ou en points) ou une valeur (diagramme en points) ?



21

UNE DONNÉE QUALITATIVE ET UNE QUANTITATIVE

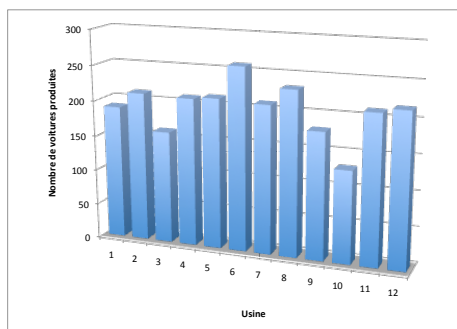
- La donnée qualitative est mis sur l'axe X. Est-elle ordonnée ou non ? Si oui, faire attention à l'ordre des catégories.
- La donnée quantitative est-elle un dénombrement (diagramme en bâtons ou en points) ou une valeur (diagramme en points) ?



22

EVITER LES FIORITURES QUI N'ONT PAS DE SIGNIFICATION

Comment interpréter le dégradé; comment interpréter la profondeur ?



23

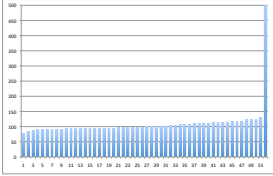
INDIQUER L'ERREUR SUR DES MESURES

- 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



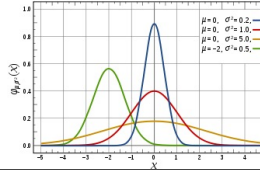
- Les maximums et minimums sont très sensibles aux valeurs extrêmes qui peuvent être aberrantes.

90,94395875
104,2067824
114,7314307
106,9794451
96,05859194
105,5617497
93,65674798
109,1262714
92,72690771
109,2090865
121,8742706
94,25174591
92,5170973
93,98157116
129,67119
114,9991892
110,2094402
96,39726327
90,72694514
97,05399126
502,0900043
99,05621255
116,6650301
89,83580927
89,31494963
111,3529959
90,44358088
116,0476333
124,7015074
101,4738367
105,2559631
98,04336577
78,2080103
92,32118279
95,31785292
99,21751704
115,2048455
93,90786374
124,0322955
90,7379407
105,7477515
98,82864594
98,13813476
113,3859338
101,4243429
94,5942288
97,59421019
94,1170698
113,5059867
85,2208859
87,27059421
94,52078999

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 $\pm 2SD$
 - Cela reste à peu près vrai si la distribution est unimodale et symétrique.



90,94395875
104,2067824
114,7314307
106,9794451
96,05859194
105,5617497
93,65674798
109,1262714
92,72690771
109,2090865
121,8742706
94,25174591
92,5170973
93,98157116
129,67119
114,9991892
110,2094402
96,39726327
90,72694514
97,05399126
502,0900043
99,05621255
116,6650301
89,83580927
89,31494963
111,3529959
90,44358088
116,0476333
124,7015074
101,4738367
105,2559631
98,04336577
78,2080103
92,32118279
95,31785292
99,21751704
115,2048455

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 $\pm 2SE$ (conclusion assez robuste à la forme de la distribution)

90,94395875
104,2067824
114,7314307
106,9794451
96,05859194
105,5617497
93,65674798
109,1262714
92,72690771
109,2090865
121,8742706
94,25174591
92,5170973
93,98157116
129,67119
114,9991892
110,2094402
96,39726327
90,72694514
97,05399126
502,0900043
99,05621255
116,6650301
89,83580927
89,31494963
111,3529959
90,44358088
116,0476333
124,7015074
101,4738367
105,2559631
98,04336577
78,2080103
92,32118279
95,31785292
99,21751704
115,2048455
93,90786374
124,0322955
90,7379407
105,7477515
98,82864594
98,13813476
113,3859338
101,4243429
94,5942288
97,59421019
94,1170698
113,5059867
85,2208859
87,27059421
94,52078999

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.

78,2080103	1,92%
85,22088859	3,85%
87,27059421	5,77%
89,31484063	7,69%
89,83580927	9,62%
90,44358088	11,54%
90,72884514	13,46%
90,73798407	15,38%
90,84385815	17,31%
92,32118279	19,23%
92,5179973	21,15%
92,72808771	23,08%
93,65674798	25,00%
93,90788374	26,92%
93,88157116	28,85%
94,11170688	30,77%
94,25174581	32,69%
94,52078889	34,62%
94,59342288	36,54%
95,31785282	38,46%
96,05858194	40,38%
96,38748327	42,31%
97,05389126	44,23%
97,59421019	46,15%
98,04336377	48,08%
98,13613476	50,00%
98,82864584	51,92%
99,05821255	53,85%
99,21761704	55,77%
101,4243429	57,69%
101,4738367	59,62%
104,2067824	61,54%
105,258631	63,46%
105,5617487	65,38%
105,7477615	67,31%
105,9794851	69,23%
109,1262714	71,15%
109,2090865	73,08%
110,2004402	75,00%
111,3528959	76,92%
113,389338	78,85%
113,5505867	80,77%
114,7314307	82,69%
116,9991892	84,62%
116,2048455	86,54%
116,0476333	88,46%
116,6650301	90,38%
121,8742706	92,31%
124,0322955	94,23%
124,7015074	96,15%
129,67118	98,08%
202,0900443	100,00%