

Chapitre 2

Les graphiques



I. Généralités

C'est la partie des statistiques la moins souvent oubliée dans l'enseignement secondaire car elle mobilise la notion de proportionnalité sous ses différentes formes.

Les graphiques sont de natures très diverses : un tableur comme Excel offre de multiples possibilités.

Ils posent cependant de nombreuses interrogations qui ont été formulées par J.C. Girard dans la revue *Repères* n° 23 (avril 1996) :

- Sur le sens des graphiques
 - Quel est l'avantage d'un graphique sur un tableau de valeurs ?
 - Le graphique sert-il d'illustration ou permet-il de découvrir une structure des données que le tableau ne mettait pas en évidence ?
 - Peut-on repasser du graphique au tableau ?
 - Quelle perception de la réalité a-t-on en regardant un graphique ?
 - Pourquoi tel graphique plutôt que tel autre ? Dans quels cas, chacun est-il pertinent ?
- Sur des notions qui renvoient à différents domaines mathématiques
 - Les camemberts utilisent la notion d'angle et de mesure d'angle qui ne sont pas toujours acquises. Comment peut-on prendre en compte cet état de fait ? Que représente le disque complet ? Autrement dit, quel est l'ensemble sur lequel on calcule les *pourcentages* ?
 - Les histogrammes et les graphiques en barres ou en bâtons utilisent une échelle verticale sur laquelle on porte des effectifs ou des *fréquences*. Sur quel ensemble de référence ces fréquences ont-elles été calculées ?
 - Lorsqu'on représente des variations, sont-elles calculées de façon absolue ou relativement à une valeur de référence ?

Exemple 1 (extrait d'un livre de CM1 : Objectif Calcul - Éditions Hatier)

Le livre pose les questions suivantes :

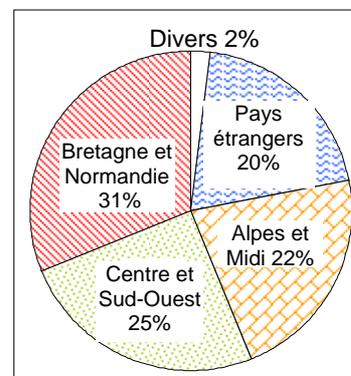
- 1) Observe ce graphique.
- 2) Essaie de le lire.
- 3) Quels renseignements donne-t-il ?
- 4) Essaie de traduire ce graphique par un tableau de nombres.

On pourrait aussi demander (en CM2, en 6^e ou plus tard !) :

- Sur quoi sont calculés les pourcentages ?
- Est-ce 20% des français qui partent en vacances

à l'étranger ou 20% de ceux qui partent en vacances qui vont à l'étranger ?

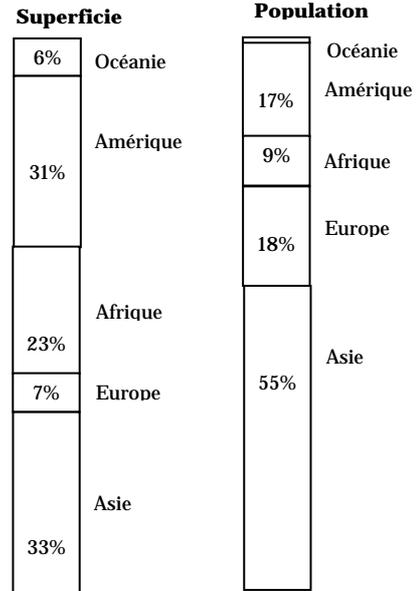
- Peut-on calculer combien de français partent à l'étranger ? combien partent en vacances ? etc.



Exemple 2

(extrait du même livre)

- Que représente la longueur de chaque barre ?
- Sur quoi ont été calculés les pourcentages ?
- Peut-on comparer ces différents pourcentages ?
- Quelle idée veut donner ce graphique ?



Superficie et population des continents

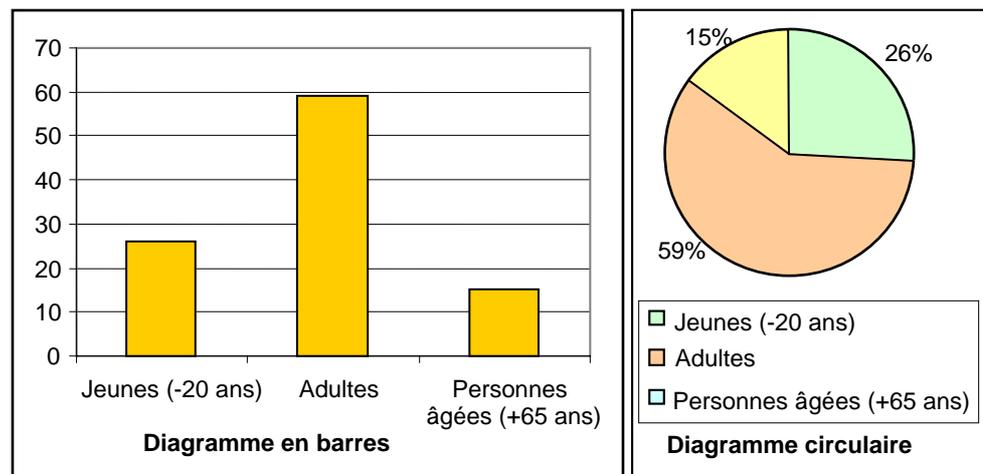
L'objectif essentiel des graphiques est de représenter la série statistique.

Comme toute représentation, ces graphiques doivent être :

- *lisibles* (les données représentées doivent pouvoir être lues),
- *fidèles* (la réalité des données ne doivent pas être déformées par la réalisation du graphique),
- *autosuffisants* (tous les renseignements doivent être mis dans la légende y compris l'ensemble de référence).

On l'a compris **chaque graphique** doit être pertinent par rapport aux données et à l'objectif poursuivi.

2. Cas d'un caractère qualitatif



Le *diagramme en barres* permet de comparer les parties entre elles. Lorsque les modalités sont ordonnées par effectifs décroissants, on obtient un diagramme dit de Pareto. La longueur de la barre est proportionnelle aux effectifs ou à la fréquence.

Le *diagramme à secteurs* (circulaire ou semi-circulaire) : il permet de comparer la partie au tout. L'aire du secteur est proportionnelle à l'effectif ou à la fréquence.

Si l'on veut comparer plusieurs diagrammes à secteurs entre eux (sur plusieurs années par exemple), les rayons doivent être proportionnels à la racine carrée de l'effectif total.

Ces deux types de représentation nécessitent d'avoir une bonne perception mentale d'un pourcentage soit dans le domaine des longueurs soit dans celui des aires.

3. Cas d'un caractère quantitatif discret

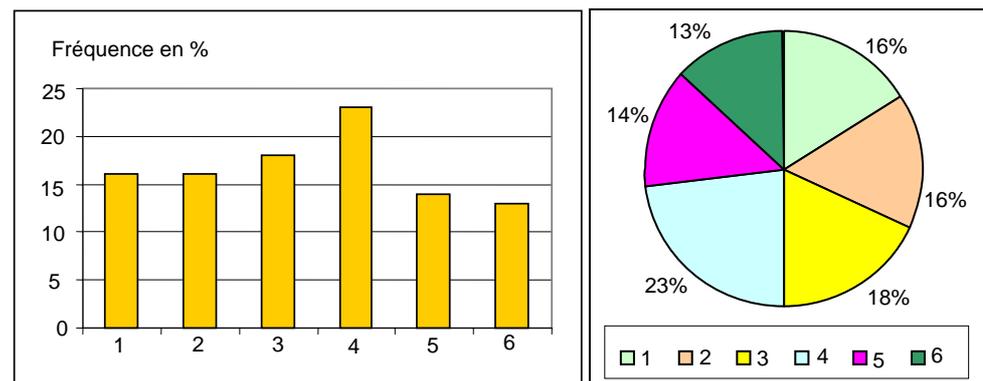
On utilisera :

Le *diagramme en bâtons* (ou à bandes) : il permet de facilement comparer les effectifs ou les fréquences entre eux.

Relier les sommets entre eux pour constituer ce qui est nommé parfois le polygone des effectifs n'a pas de sens : en effet les points des segments autres que les extrémités n'ont aucune signification !

On peut également utiliser un *diagramme à secteurs* si on souhaite comparer la partie au tout. Toutefois le diagramme semi-circulaire doit être privilégié pour respecter une structure d'ordre à mettre en évidence dans les modalités.

Exemple : Série de 100 lancers d'un dé équilibré



4. Cas d'un caractère quantitatif continu

4.1. On utilise un histogramme

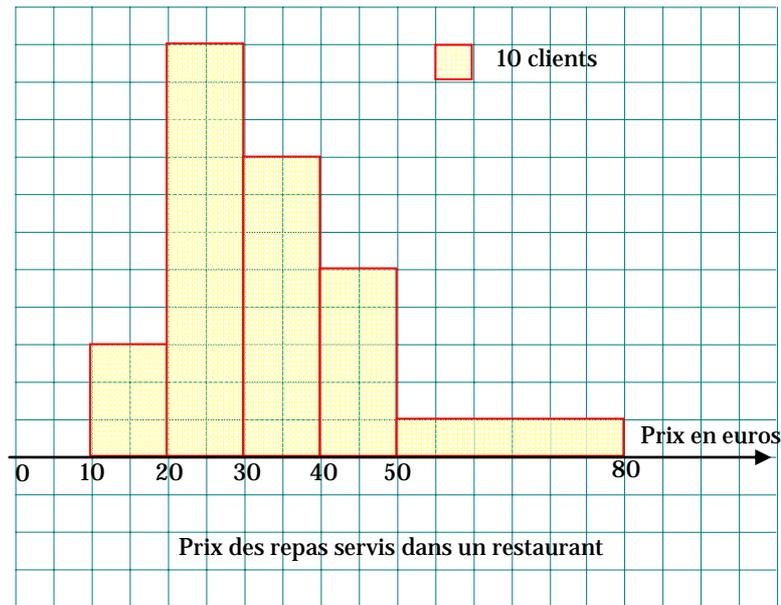
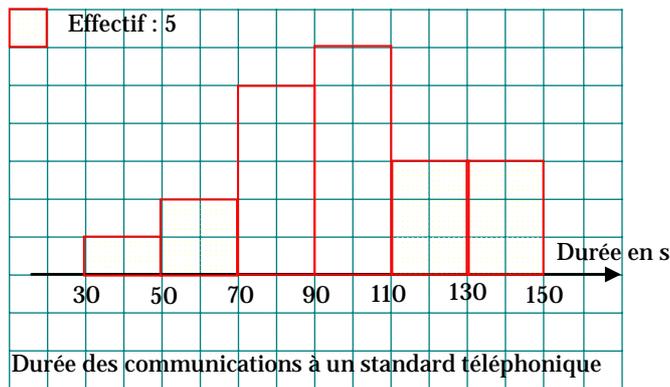
Il est constitué par des rectangles contigus ayant pour base chacune des classes et une aire proportionnelle à l'effectif ou à la fréquence de la classe correspondante.

Si les amplitudes sont toutes égales, la hauteur du rectangle est proportionnelle à l'effectif (ou à la fréquence).

Si les amplitudes sont inégales, la hauteur est proportionnelle à la densité de la classe.

En théorie, l'histogramme est la représentation graphique de la densité en tant que fonction des différents intervalles de la partition de l'ensemble des modalités.

En pratique, on indique une unité d'aire correspondante à un certain effectif (ou à une certaine fréquence).



4.2. Sur la notion d'histogramme

On a trop tendance à considérer l'histogramme comme une juxtaposition de rectangles dont l'intérêt se limite :

- à l'obtention d'un dessin,
- à satisfaire un item du programme.

Prenons l'exemple de la modélisation du contrôle du réglage d'une machine fabriquant des profilés aluminium.

Il paraît pertinent de choisir une **variable quantitative continue** comme la longueur X en cm des profilés. Il est non moins pertinent de choisir cette variable **X absolument continue**.

RAPPEL

Une variable X est absolument continue s'il existe une fonction f définie sur \mathbb{R} telle que :

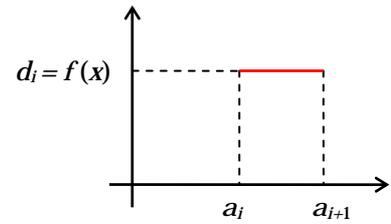
- f est positive sur \mathbb{R} ,
- f est continue sur \mathbb{R} sauf peut-être en un nombre fini de points où elle admet une limite à droite et une limite à gauche,

- $\int_{-\infty}^{+\infty} f(t) dt = 1$,

- La fonction de répartition F de X est liée à f par : $F(x) = \int_{-\infty}^x f(t) dt$.

On dit que f est une **densité** de X . Abusivement, f est appelée **loi de X** .

Les données peuvent être placées dans des intervalles dont l'amplitude correspond à la précision de la machine (le mm par exemple). On choisit un nombre fini d'intervalles notés $[a_i; a_{i+1}[$ pour $1 \leq i \leq k$.



A priori, tous les nombres réels appartenant à un tel intervalle ont la même chance d'être le résultat d'une mesure.

Cela se traduit pour le modèle adopté par une densité (de fréquence) constante sur chaque intervalle.

Pour estimer les valeurs d_i , on procède alors à un sondage en mesurant le plus grand nombre possible de profils dans des conditions admises identiques. On obtient la série statistique des mesures qui sont réparties entre les différents intervalles :

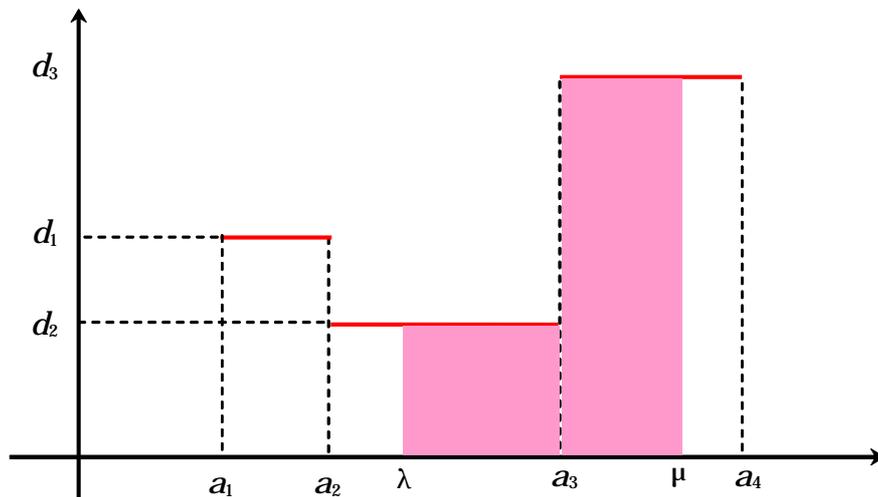
Intervalles	$]-\infty; a_1[$	$[a_1; a_2[$...	$[a_i; a_{i+1}[$...	$[a_k; a_{k+1}[$	$[a_{k+1}; +\infty[$
Effectifs	0	n_1	...	n_i	...	n_k	0
Fréquences	0	f_1	...	f_i	...	f_k	0

On a alors : $d_i = \frac{f_i}{a_{i+1} - a_i}$.

REMARQUE

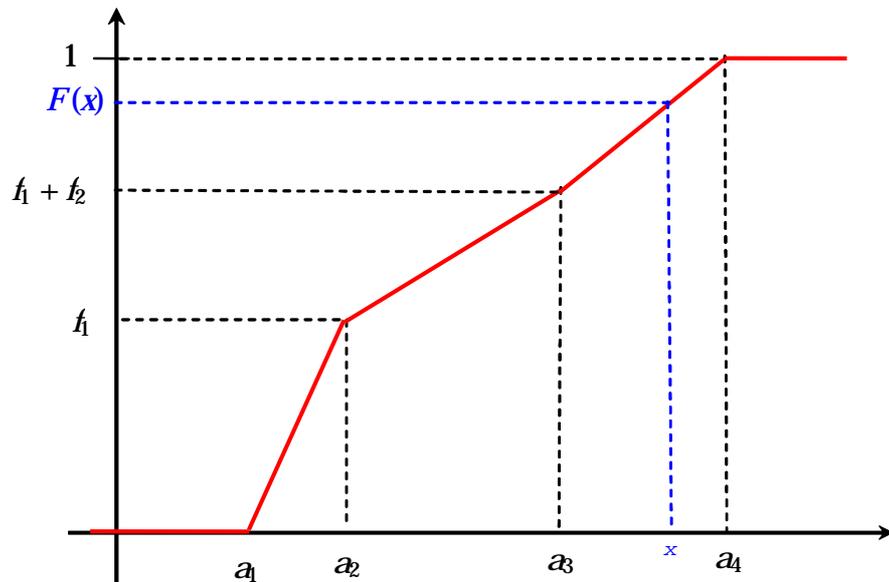
L'unité sur l'axe des ordonnées est alors : % / cm . Il est donc erroné d'interpréter l'axe des ordonnées comme axe des effectifs ou des fréquences, ce qui se fait malheureusement souvent.

La fonction densité, obtenue par observation statistique et représentée par l'histogramme, est une fonction constante par intervalles :

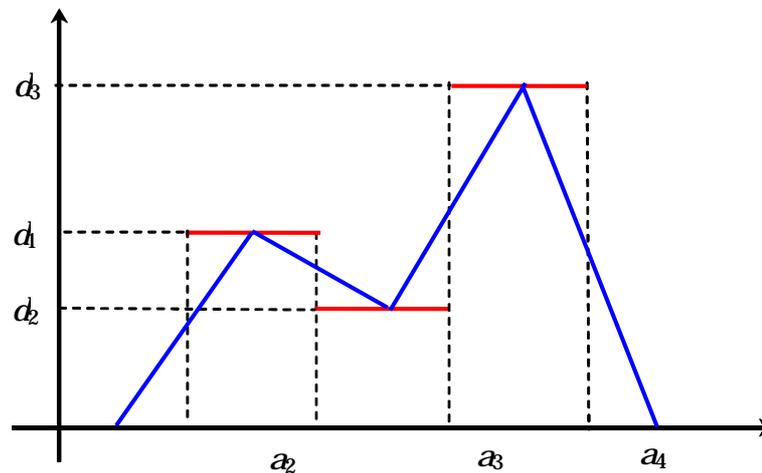


La fréquence possible des profils dont la longueur est comprise entre λ et μ est estimée par l'aire du domaine rosé.

On pourrait aussi estimer la fréquence possible $F(x)$ des profils dont la longueur est inférieure ou égale à x . Cette fonction F est la **fonction cumulative croissante** ou **fonction de répartition** de X . C'est une primitive de la densité f . La courbe obtenue est une fonction continue et affine par intervalle.



Dans de nombreux ouvrages sur la statistique, on trouve la notion de *polygone des effectifs ou des fréquences*. Dans le cas où les classes ont même amplitude, on suggère de tracer sur l'histogramme une ligne brisée reliant les milieux des côtés supérieurs des rectangles de chaque classe. Lorsque les classes n'ont pas la même amplitude, on se ramène au cas précédent par un découpage *ad-hoc* des classes.



Il s'agit de remplacer l'histogramme précédent par un autre histogramme représenté par une courbe limitant un domaine d'aire égale à 1 (ou à l'effectif total) et **continu**. L'idée générale est de lisser la courbe de densité pour la comparer à celle de variables connues servant de modèles et le procédé suggéré précédemment fournit **une** solution à ce problème.

Pour conclure, on peut retenir cette définition proposée par J.C. Régnier :

Soit X une variable statistique (resp. aléatoire) absolument continue de densité f . L'histogramme est la surface limitée par la représentation graphique de f et l'axe des abscisses.

5. Graphiques en tiges et feuilles (stem and leaf)

La réalisation d'un histogramme n'est pas exempte de difficultés pour les élèves : notion d'intervalle semi-ouvert, utilisation d'une échelle, proportionnalité, concept de densité...

Le graphique tiges-feuilles (John W. Tukey) est assez proche de l'histogramme sans avoir les obstacles précédemment cités.

Prenons l'exemple cité par J.C. Girard : les données suivantes représentent la dureté, en indice Rockwell, de 60 pièces mécaniques après trempage.

Dans un premier temps les valeurs sont arrondies à l'unité pour pouvoir travailler sur des entiers.

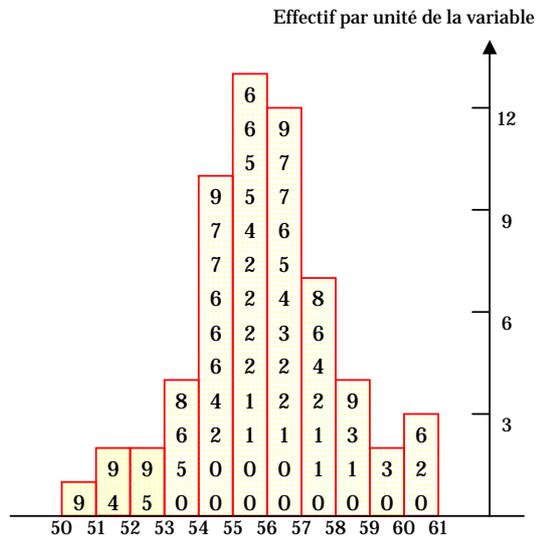
51	54	55	56	57	58	Effectif	
51	54	55	56	57	58	51	2 ★★
52	54	55	56	57	58	52	1 ★
53	55	55	56	57	58	53	3 ★★★
53	55	55	56	57	58	54	7 ★★★★★★★
53	55	55	56	57	59	55	15 ★★★★★★★★★★★★
53	55	55	56	57	59	56	11 ★★★★★★★★★★★
54	55	55	56	57	59	57	10 ★★★★★★★★★★★
54	55	55	56	57	60	58	5 ★★★★★
54	55	56	56	57	60	59	3 ★★★
54	55	56	57	58	61	60	2 ★★
						61	1 ★

On travaille maintenant sur les valeurs au dixième : on obtient 44 valeurs différentes. Un schéma analogue au précédent ne présente guère d'intérêt. Le graphique « Stem and leaf » constitue un préalable intéressant avant d'aborder la notion d'histogramme.

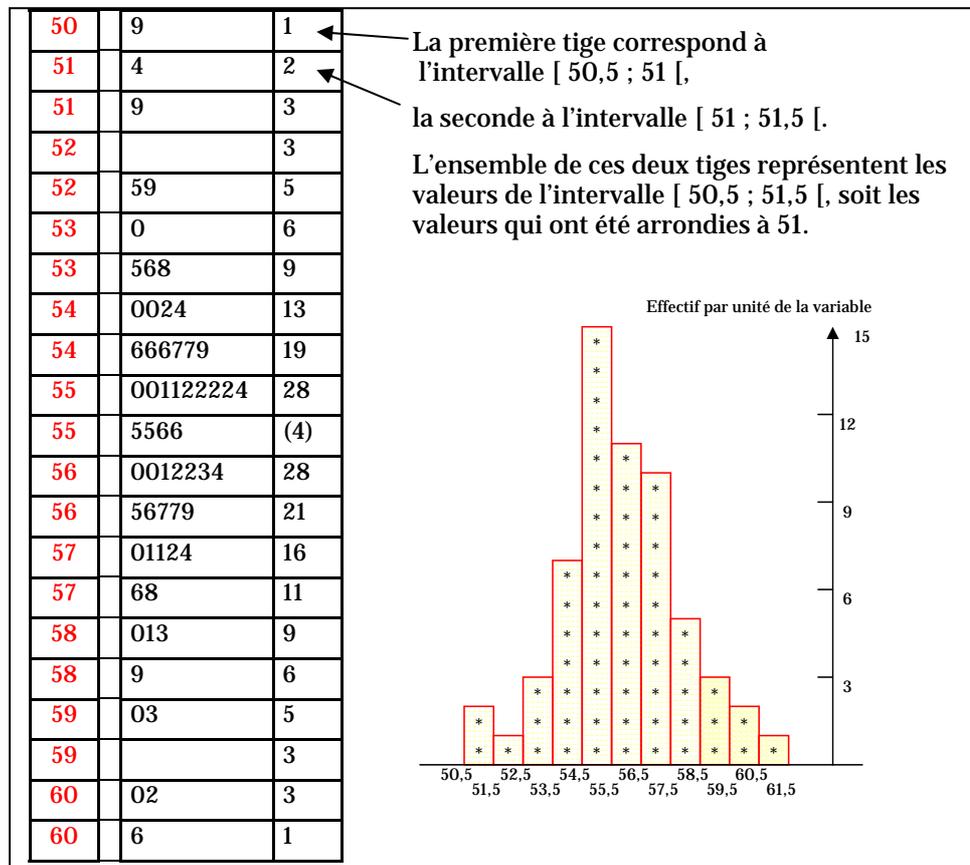
50,9	54,0	55,0	55,6	56,6	57,8	50	9	1
51,4	54,2	55,1	55,6	56,7	58,0	51	49	3
51,9	54,4	55,1	56,0	56,7	58,1	52	59	5
52,5	54,6	55,2	56,0	56,9	58,3	53	0568	9
52,9	54,6	55,2	56,1	57,0	58,9	54	0024666779	19
53,0	54,6	55,2	56,2	57,1	59,0	55	0011222245566	(13)
53,5	54,7	55,2	56,2	57,1	59,3	56	001223456779	28
53,6	54,7	55,4	56,3	57,2	60,0	57	0112468	16
53,8	54,9	55,5	56,4	57,4	60,2	58	0139	9
54,0	55,0	55,5	56,5	57,6	60,6	59	03	5
						60	026	3

Tige 52, feuilles 5 et 9

La simplicité de ce graphique est évidente, il revient à regrouper les valeurs dans des intervalles d'amplitude 1 ($[50; 51[$, $[51; 52[$, ...). Par rapport à un histogramme, il a l'avantage de ne perdre aucune information sur les données de départ. On n'est cependant pas loin de l'histogramme :



A partir des données précédentes, on peut aussi construire deux tiges par unité :



On peut alors élaborer un histogramme dont les classes ont une amplitude d'une unité.

