

Les exemples de ce document sont tirés de l'excellent livre "Exploratory data analysis" (1977), de John W. Tukey. Tukey, celui-la même qui a donné son nom au diagramme de Tukey (qu'on appelle plus communément diagramme en boîte à moustaches, c'est le diagramme qu'on verra à la fin de ce chapitre).

**Exercice 1** (page 6–9 du livre de Tukey)

Démarrons avec des publicités pour voitures d'occasion tirées du *Sunday Standard-Times* de New Bedford, Massachusetts, daté du 18 août 1968. Dans ces 18 publicités, on peut trouver entre autres des prix pour des Chevrolets d'occasion. Voici les prix demandés dans ces publicités (en dollars) :

250; 150; 795; 895; 695; 1699; 1499; 1099; 1693; 1166; 688; 1333; 895; 1775; 895; 1895; 795

Voici à quoi ressemblerait un diagramme en tiges et feuilles pour cet ensemble de données, si on choisit d'organiser nos tiges par centaines de dollars (dans cet exemple, le fait de marquer deux étoiles montre qu'on a coupé par centaines, et qu'il reste donc deux chiffres à lire dans la colonne des feuilles, pour chaque prix ; on n'est pas obligés de mettre des étoiles partout, Tukey a choisi d'en mettre à plusieurs endroits pour que ça soit plus clair, sans surcharger l'écriture) :

1**	50
2	50
3	
4	
5**	
6	95, 88
7	95, 95
8	95, 95, 95
9**	
10	99
11	66
12	
13**	33
14	99
15	
16	99, 93
17**	75
18	95

Expliquez avec des mots ce que représente la ligne du tableau qui correspond à la tige 16.

## Exercice 2 (page 14 du livre)

Parfois, les données ne sont pas adaptées pour que toutes les tiges soient similaires. Dans l'exemple à la page 9 du diaporama ([http://www.barsamian.am/2021-2022/S4P6/Chap5\\_diaporama.pdf](http://www.barsamian.am/2021-2022/S4P6/Chap5_diaporama.pdf)), une tige par heure était très adapté. Dans l'exemple de l'exercice 1, une tige par centaine de dollars était adapté. Cela se voit sur le diagramme qui est régulièrement rempli.

Parfois, et même souvent, les données ne sont pas régulièrement espacées. On a alors parfois recours à un diagramme en tiges et feuilles où les tiges n'ont pas toutes la même signification. En exemple, voici une liste de puissances de centrales hydroélectriques des États-Unis (en megawatts) :

100; 1345; 50; 162; 150; 36; 250; 114; 24; 1974; 40; 45; 285; 33; 15; 75; 225;

30; 16; 379; 120; 32; 108; 22; 25; 48; 24; 36; 900; 134; 120; 71; 424; 60

Pour ces données, Tukey choisit de montrer d'abord les petits nombres (entre 0 et 99) avec une première partie où les tiges représentent des dizaines, puis les nombres moyens (entre 100 et 999) avec une seconde partie où les tiges représentent des centaines, puis le reste avec une dernière partie où les tiges représentent des milliers :

1*		6, 5
2		5, 2, 4, 4
3		0, 3, 6, 2, 6
4		5, 0, 8
5*		0
6		0
7		5, 1
8		
9*		
1**		20, 62, 34, 50, 00, 20, 14, 08
2		25, 85, 50
3		79
4		24
5**		
6		
7		
8		
9**		00
1***		345, 974
2		
3		

Représenter ces mêmes données avec un diagramme en tiges et feuilles où les tiges sont uniquement des centaines. Pourquoi a-t-on une moins bonne idée des valeurs contenues dans la série avec cette seconde représentation ?

## Exercice 3

Faire l'exercice 6 de la feuille d'exercices distribuée en classe :

[http://www.barsamian.am/2021-2022/S4P6/Chap5\\_Exos.pdf](http://www.barsamian.am/2021-2022/S4P6/Chap5_Exos.pdf)

Indication : commencer par chercher une bonne manière d'organiser les tiges. Ici, les données proviennent des États-Unis, donc la virgule représente le séparateur de 3 chiffres, et pas le séparateur décimal (vous l'aviez remarqué car vous vous êtes bien doutés que le point culminant des États-Unis n'est pas à 20,2 pieds d'altitude, mais bien à 20 237 pieds d'altitude).