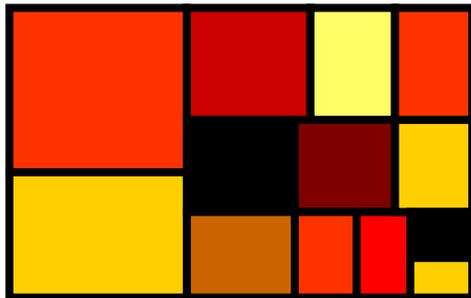
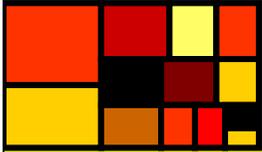


Représentations graphiques



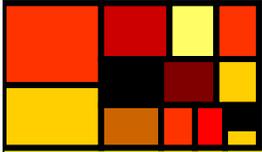
- I. Ce qu'elles signifient
- II. Exemples de représentations illicites

Types de variables



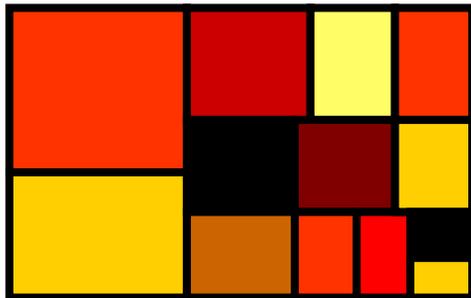
- Variables numériques :
 - Variables quantitatives continues
 - taille{1.87, 1.63, 1.893}
 - Variables quantitatives discrètes
 - nombreEnfants{0, 3, 1, 1, 2, 0, 4}
 - Rang
 - ordreNaissanceEnfant{1, 1, 5, 2, 1, 3}
- Variables qualitatives
 - Ordonnées
 - appreciation{« mauvais », « moyen », « bon », « très bon »}
 - Sans ordre
 - couleurCheveux{« blond », « brun », « gris »}
- Variables binaires : entre les deux
- Dates : variables quantitatives particulières (transformation qualitative ou non, opérations limitées)

Présentation des données (5700 lignes)



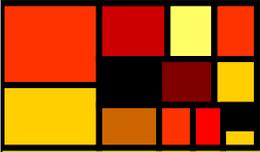
age	nb_annees_etudes	revenus	assurance	paco2	sexe	type
70.2509766	12	Under \$11k	Medicare	40	Male	white
78.1789551	12	Under \$11k	Private & Medicare	34	Female	white
46.09198	14	\$25-\$50k	Private	16	Female	white
75.3319702	9	\$11-\$25k	Private & Medicare	30	Female	white
67.9099731	10	Under \$11k	Medicare	17	Male	white
86.0779419	8	Under \$11k	Medicare	68	Female	white
54.9679871	14	\$25-\$50k	Private	45	Male	white
43.6389771	12	\$25-\$50k	Private	26	Male	white
18.0419922	13	Under \$11k	Private	40	Female	white
48.4239807	11	Under \$11k	Medicaid	30	Female	white
34.4419861	15	\$25-\$50k	Private	37	Male	white
68.3479614	12	\$11-\$25k	Private & Medicare	33	Male	white
74.7099609	12	\$25-\$50k	Private & Medicare	33	Male	white
42.2369995	16	\$25-\$50k	Private	25	Female	white
81.9709473	3	Under \$11k	Medicare	67	Male	white
78.3049927	8	Under \$11k	Medicare	37	Male	black
88.4219971	14	Under \$11k	Private & Medicare	28	Male	white
69.0019531	11	Under \$11k	Medicare	36	Female	white
...

Représentations univariées

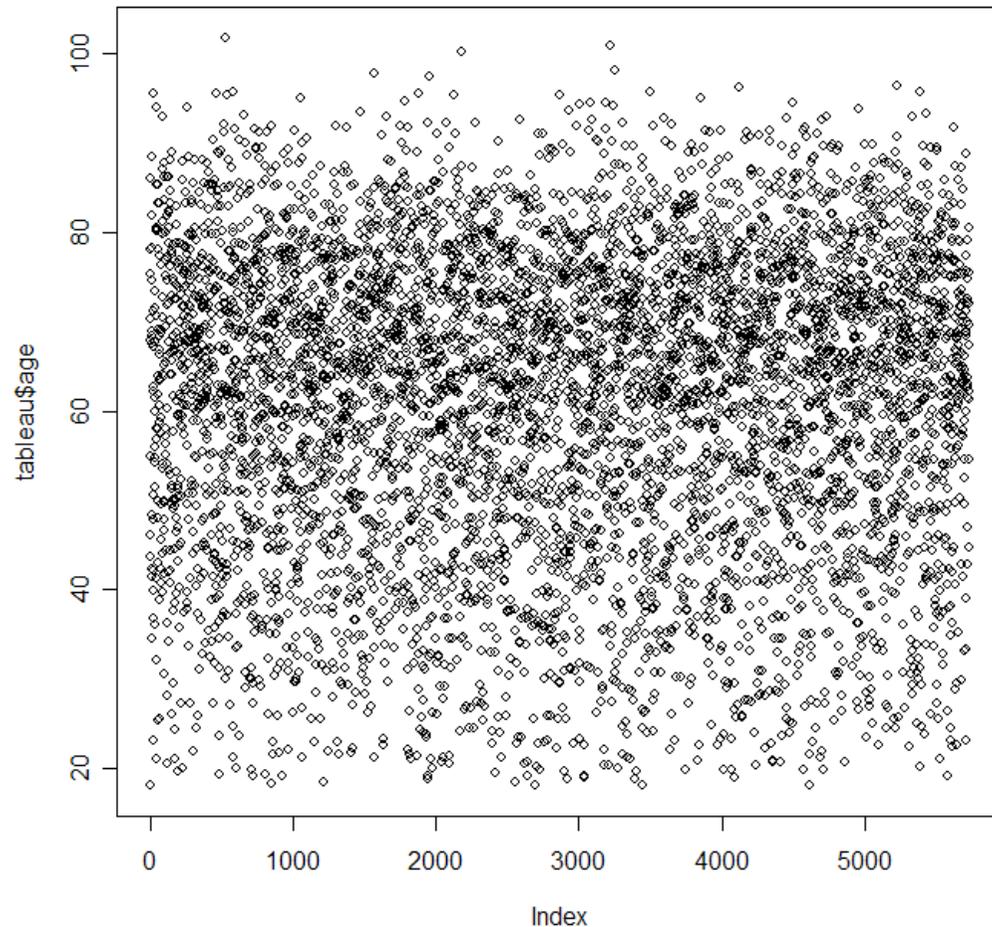


- I. Variables quantitatives continues
- II. Variables quantitatives discrètes
- III. Variables qualitatives

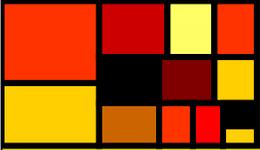
Variable quantitative continue : Représentation native



```
> plot( tableau$age );
```

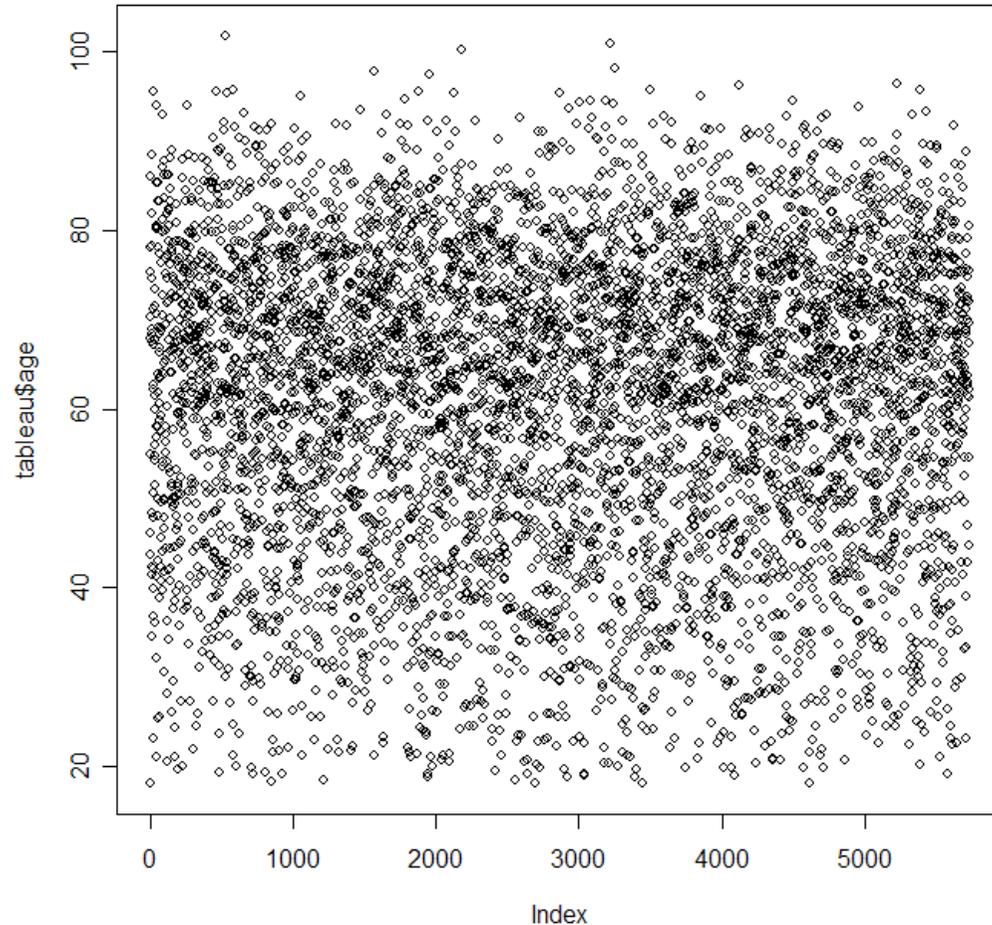


Variable quantitative continue : Représentation native

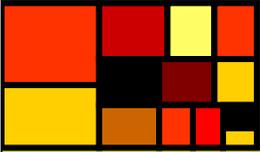


```
> plot( tableau$age );
```

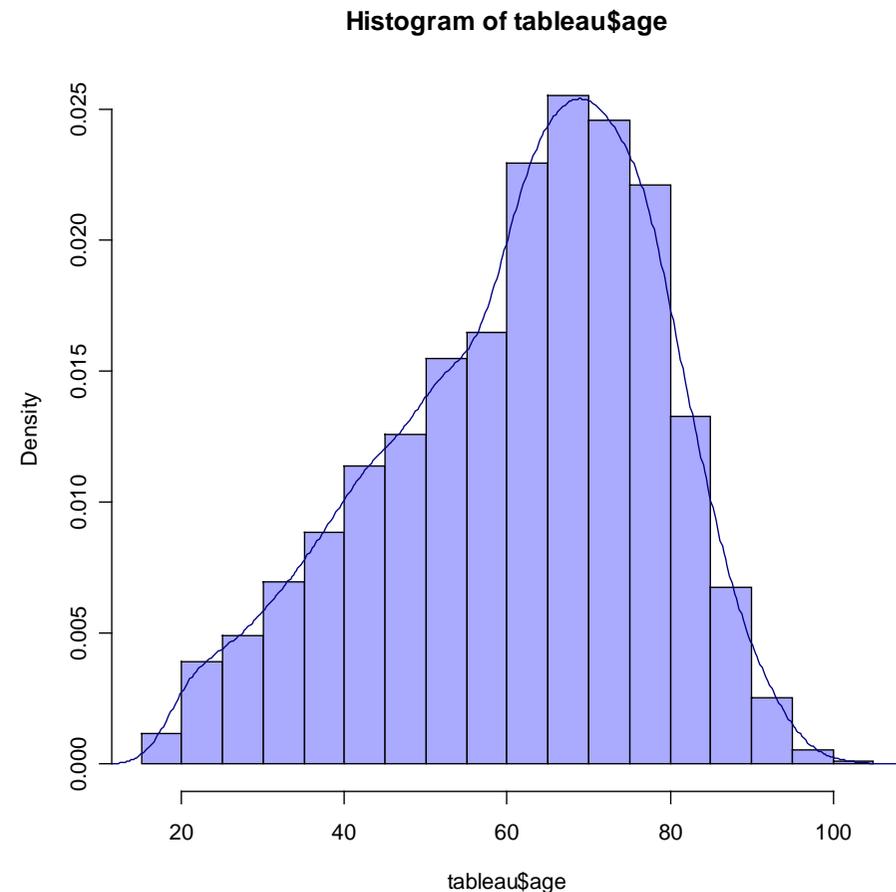
Chaque individu est une unité d'abscisse. En réalité peu importe la représentation native des données.



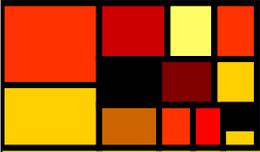
Variable quantitative continue : Représentation de la distribution par Histogramme



```
> hist(tableau$age, col="#AAAAFF", freq=FALSE) ;  
> lines(density(tableau$age), col="navy") ;
```



Variable quantitative continue : Représentation de la distribution par Histogramme

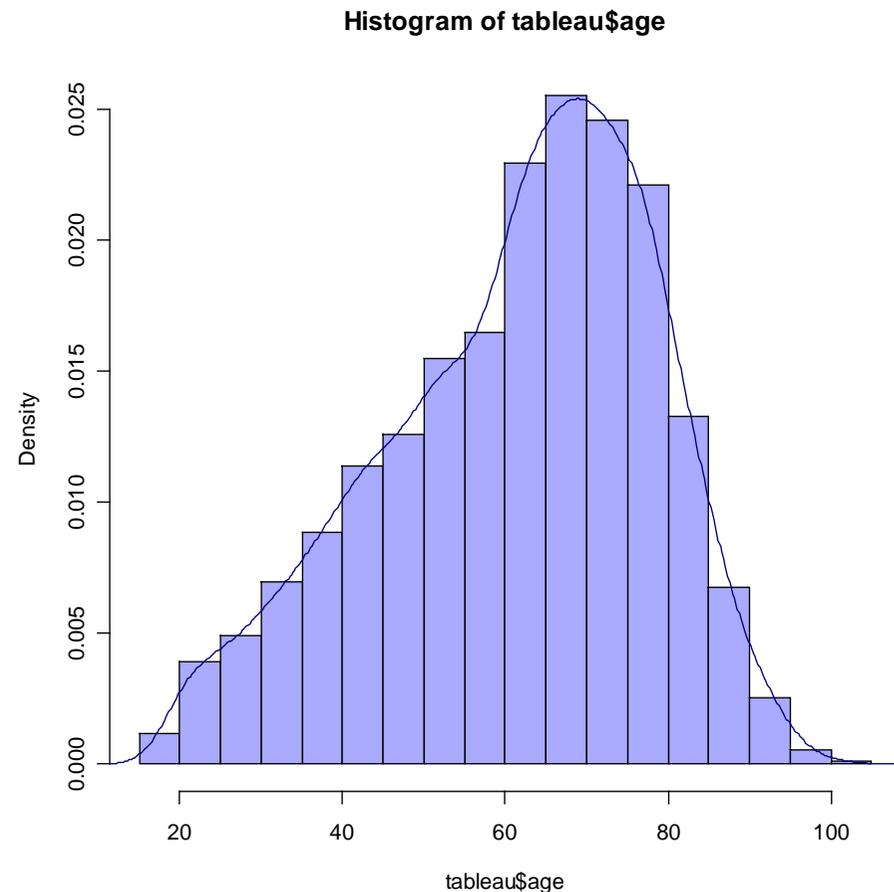


```
> hist(tableau$age, col="#AAAAFF", freq=FALSE) ;
```

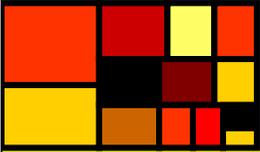
```
> lines(density(tableau$age), col="navy") ;
```

Remarques importantes :

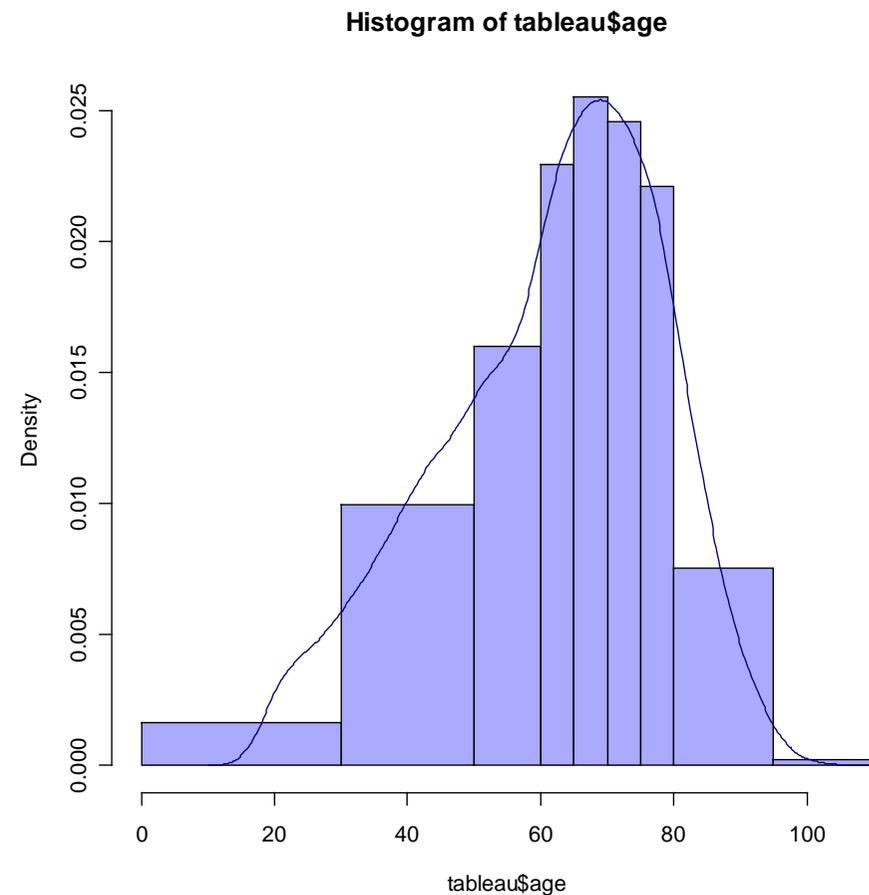
- approximation par classe de la « densité de probabilité »
- donc aire sous la courbe=1
- les hauteurs des barres ne sont pas interprétables
- en abscisse ce ne sont pas des catégories mais une véritable variable quantitative
- attention, Excel ne fait pas des histogrammes mais des diagrammes en bâtons.



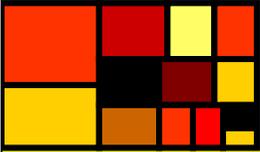
Variable quantitative continue : Représentation de la distribution par Histogramme



```
> hist(tableau$age, col="#AAAAFF", freq=FALSE,  
       breaks=c(0,30,50,60,65,70,75,80,95,110)) ;  
> lines(density(tableau$age), col="navy") ;
```



Variable quantitative continue : Représentation de la distribution par Histogramme

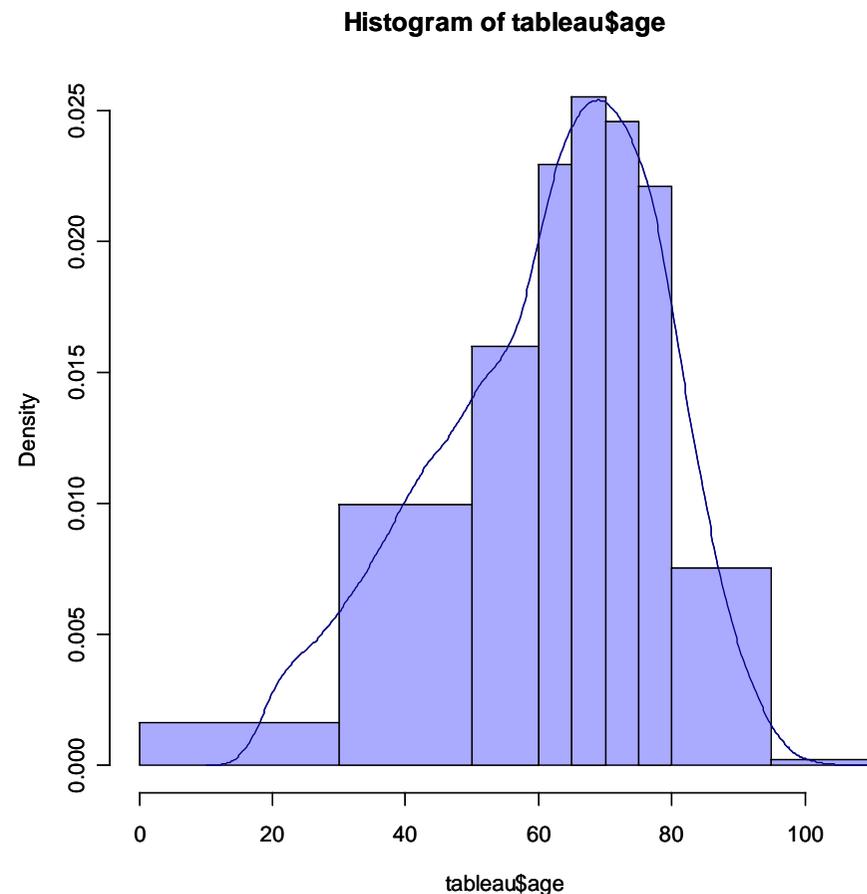


```
> hist(tableau$age, col="#AAAAFF", freq=FALSE,  
       breaks=c(0,30,50,60,65,70,75,80,95,110)) ;  
> lines(density(tableau$age), col="navy") ;
```

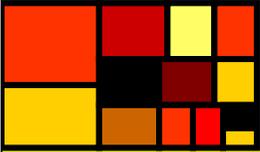
Dans l'histogramme on peut imposer les classes.

Il n'est pas correct d'utiliser la hauteur des rectangles pour représenter le nombre d'individus. Si c'était le cas, en fusionnant les barres, leur surface augmenterait.

En utilisant la surface des rectangles et non leur hauteur, on réalise bien une approximation de la densité de probabilité.



Variable quantitative continue : Représentation de la distribution par Histogramme

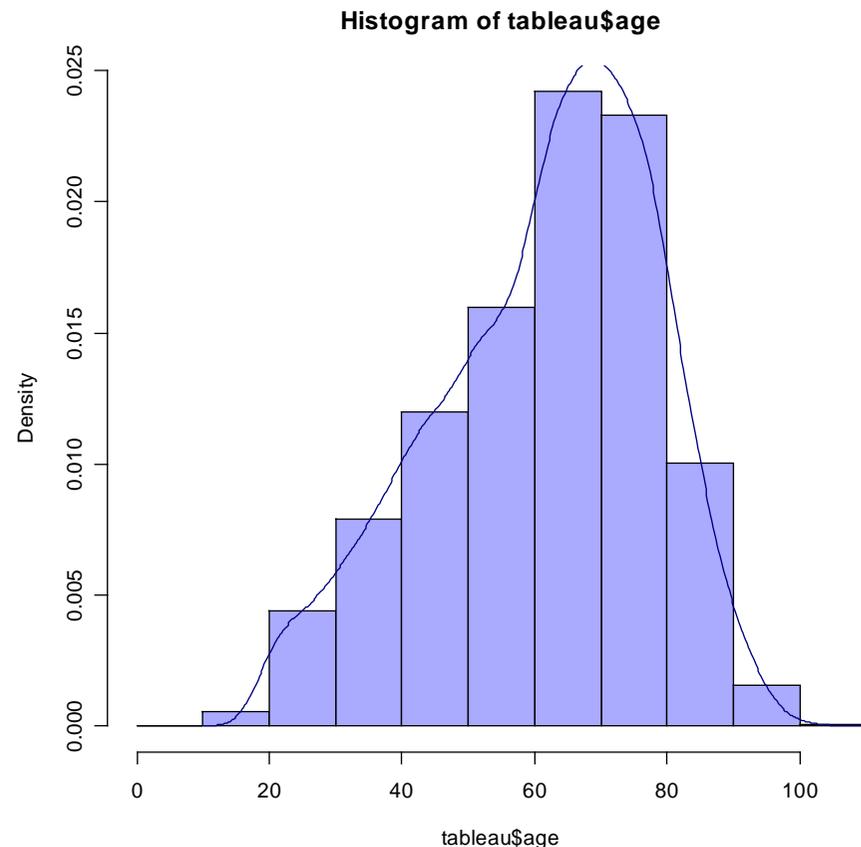


```
> hist(tableau$age, col="#AAAAFF", freq=FALSE,  
       breaks=c(0,30,50,60,65,70,75,80,95,110)) ;  
> lines(density(tableau$age), col="navy") ;
```

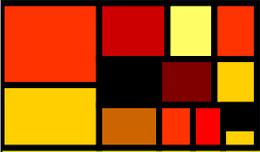
Dans l'histogramme on peut imposer les classes.

Il n'est pas correct d'utiliser la hauteur des rectangles pour représenter le nombre d'individus. Si c'était le cas, en fusionnant les barres, leur surface augmenterait.

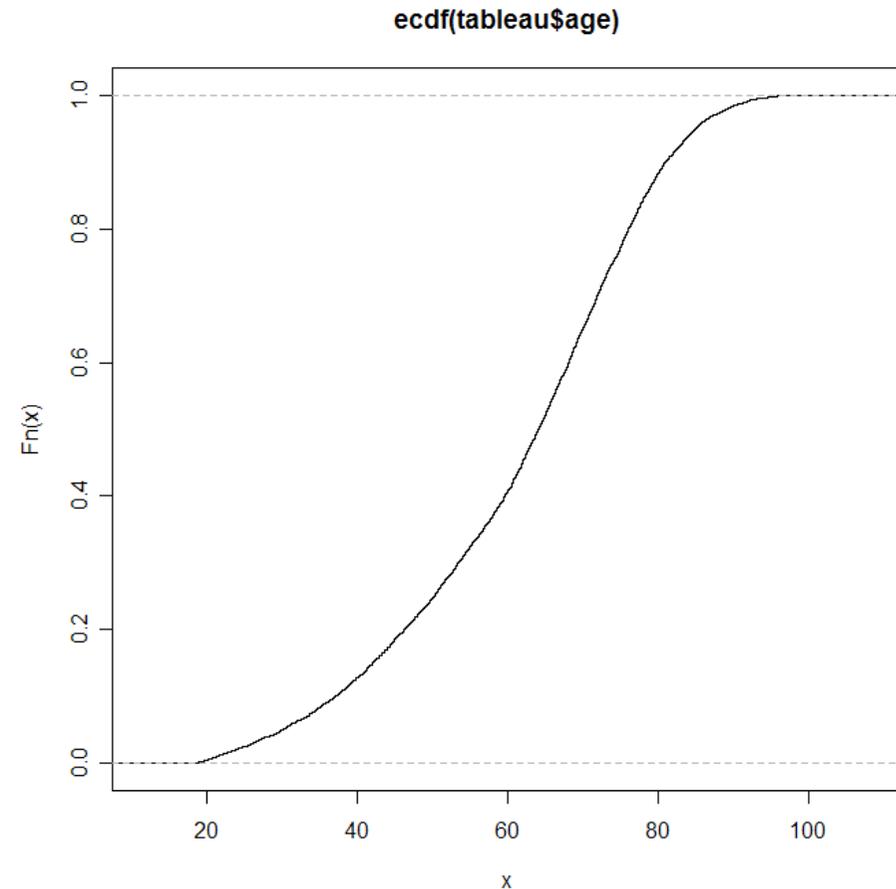
En utilisant la surface des rectangles et non leur hauteur, on réalise bien une approximation de la densité de probabilité.



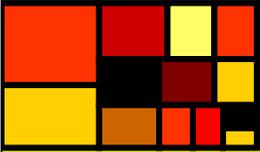
Variable quantitative continue : Représentation de la distribution par Fonction de répartition



```
> plot(ecdf(tableau$age), verticals=TRUE,  
      do.points=FALSE) ;
```

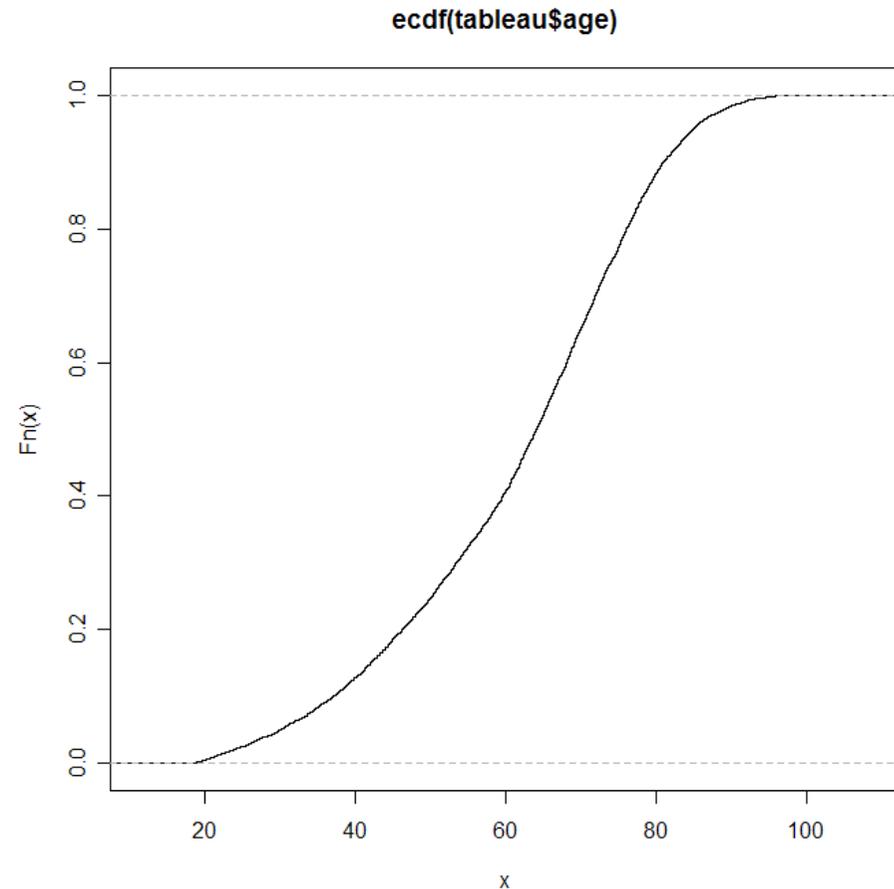


Variable quantitative continue : Représentation de la distribution par Fonction de répartition

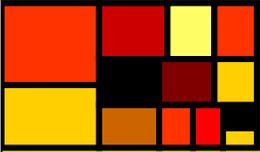


```
> plot(ecdf(tableau$age), verticals=TRUE,  
      do.points=FALSE);
```

ECDF : empirical cumulative
distribution function
C'est une estimation de la primitive de
la densité de probabilité. Elle peut être
facilement refaite avec Excel.
Toujours autorisée, elle est
intéressante surtout pour les
superpositions.
Elle fournit directement la médiane.

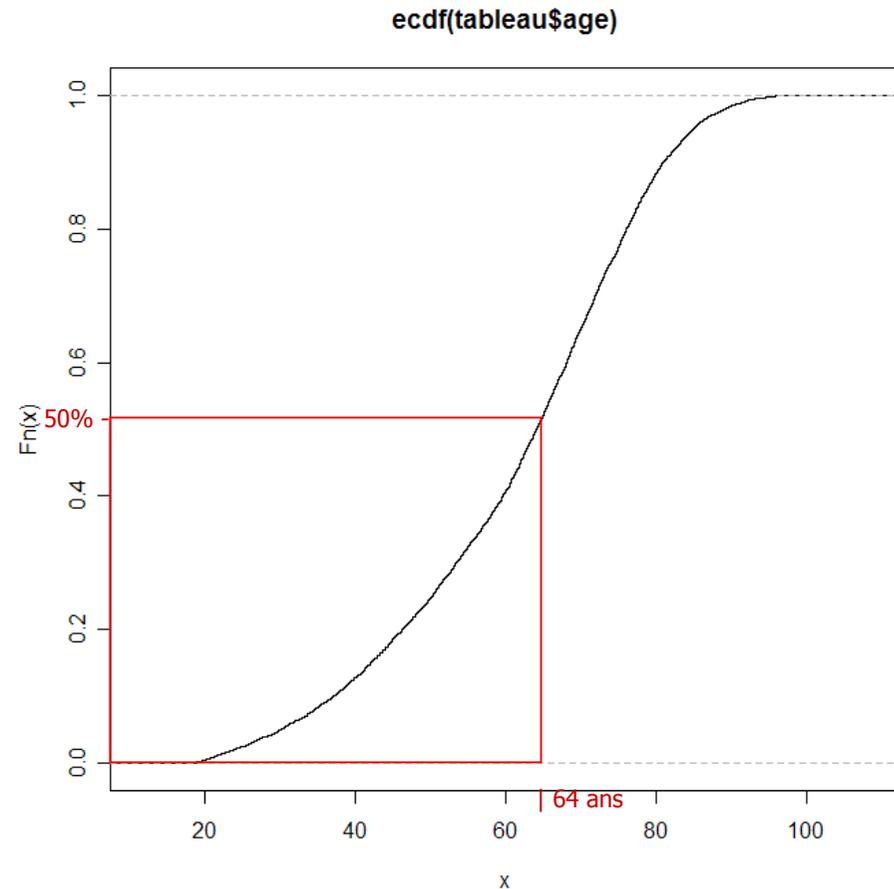


Variable quantitative continue : Représentation de la distribution par Fonction de répartition

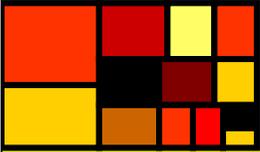


```
> plot(ecdf(tableau$age), verticals=TRUE,  
      do.points=FALSE);
```

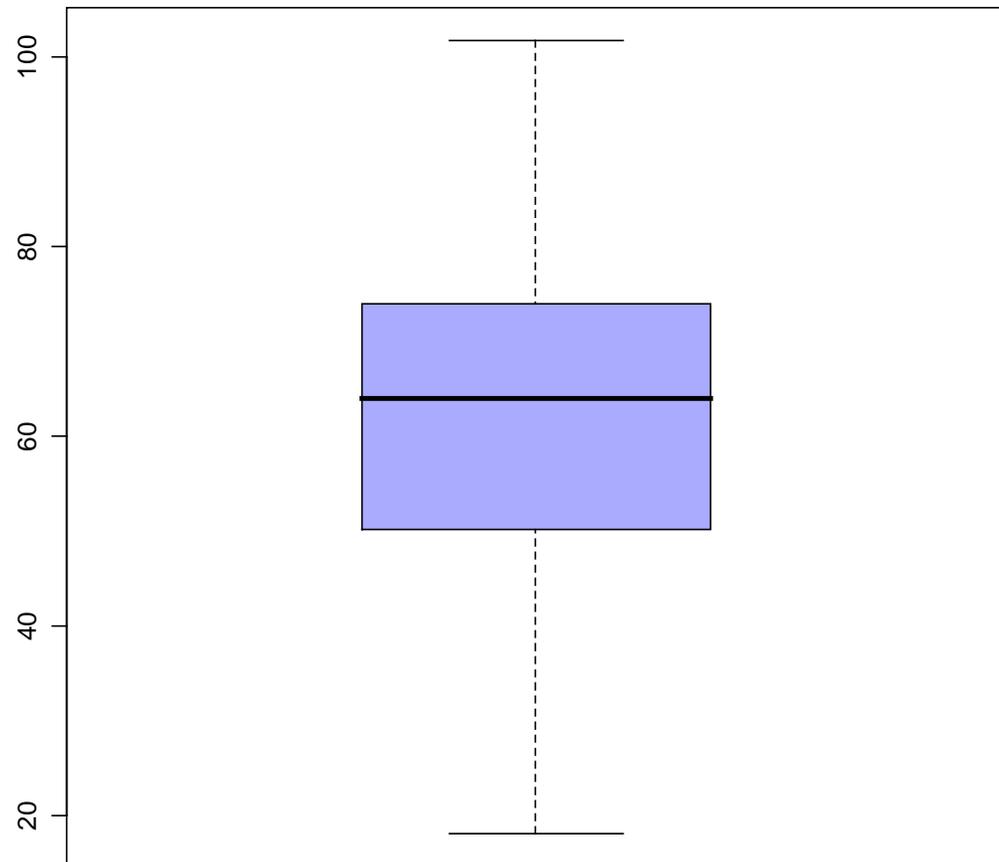
ECDF : empirical cumulative
distribution function
C'est une estimation de la primitive de
la densité de probabilité. Elle peut être
facilement refaite avec Excel.
Toujours autorisée, elle est
intéressante surtout pour les
superpositions.
Elle fournit directement la médiane.



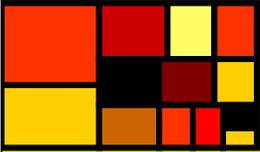
Variable quantitative continue : Représentation de la distribution par Boxplot



```
> boxplot(tableau$age, col="#AAAAFF");
```

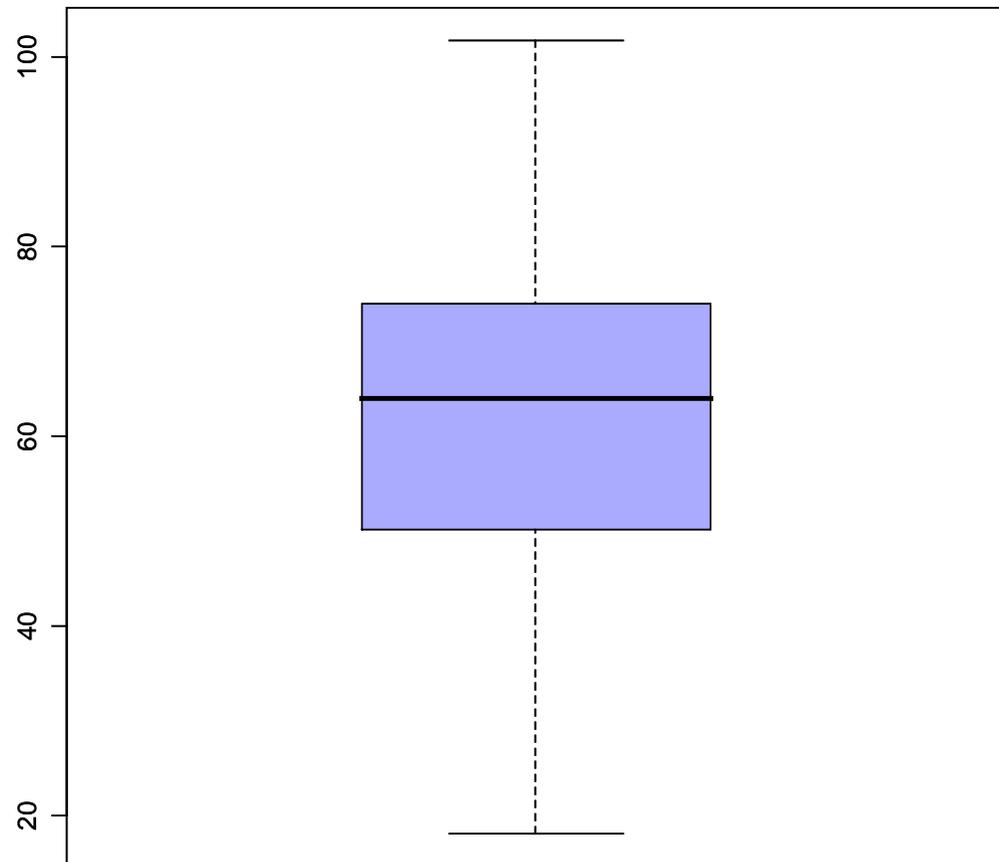


Variable quantitative continue : Représentation de la distribution par Boxplot

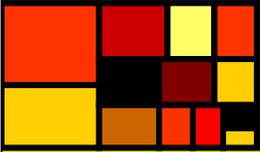


```
> boxplot(tableau$age, col="#AAAAFF");
```

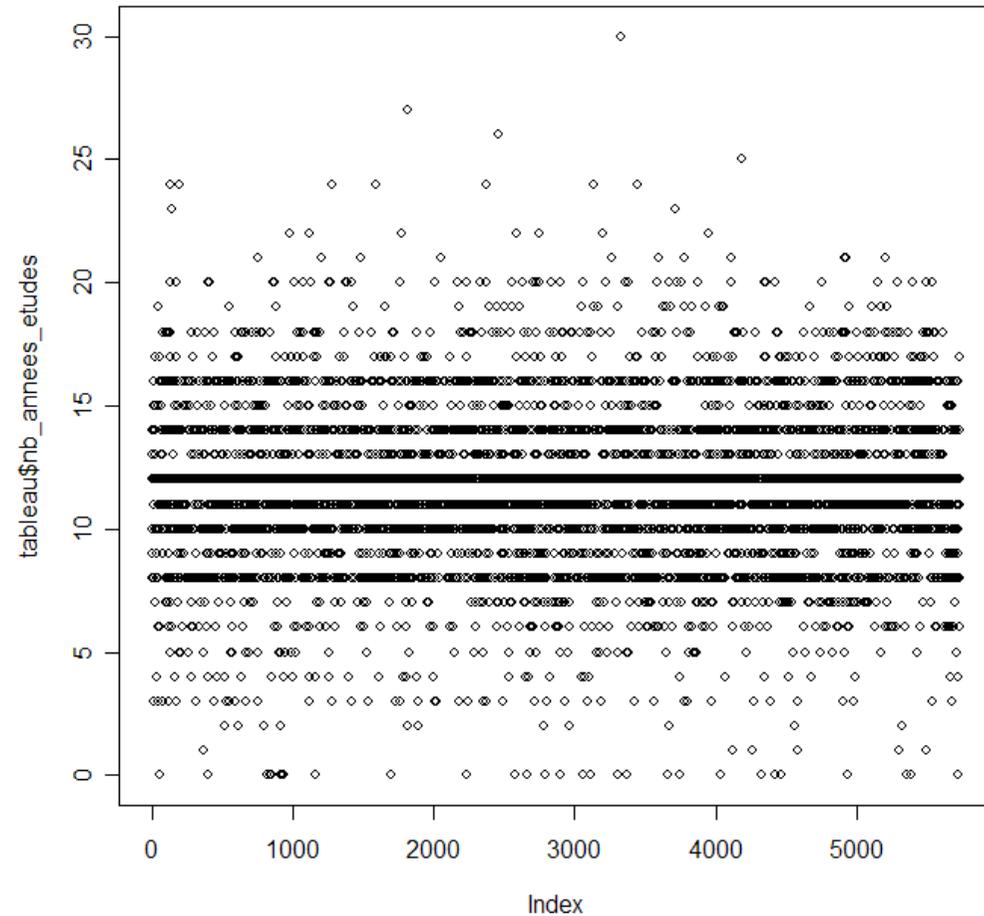
La boxplot (ou boîte à moustaches) représente des statistiques agrégées (médiane, quartiles...). La variable quantitative apparaît en ordonnées.



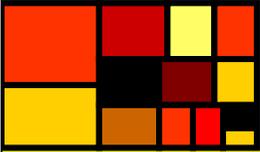
Variable quantitative discrète : Représentation native



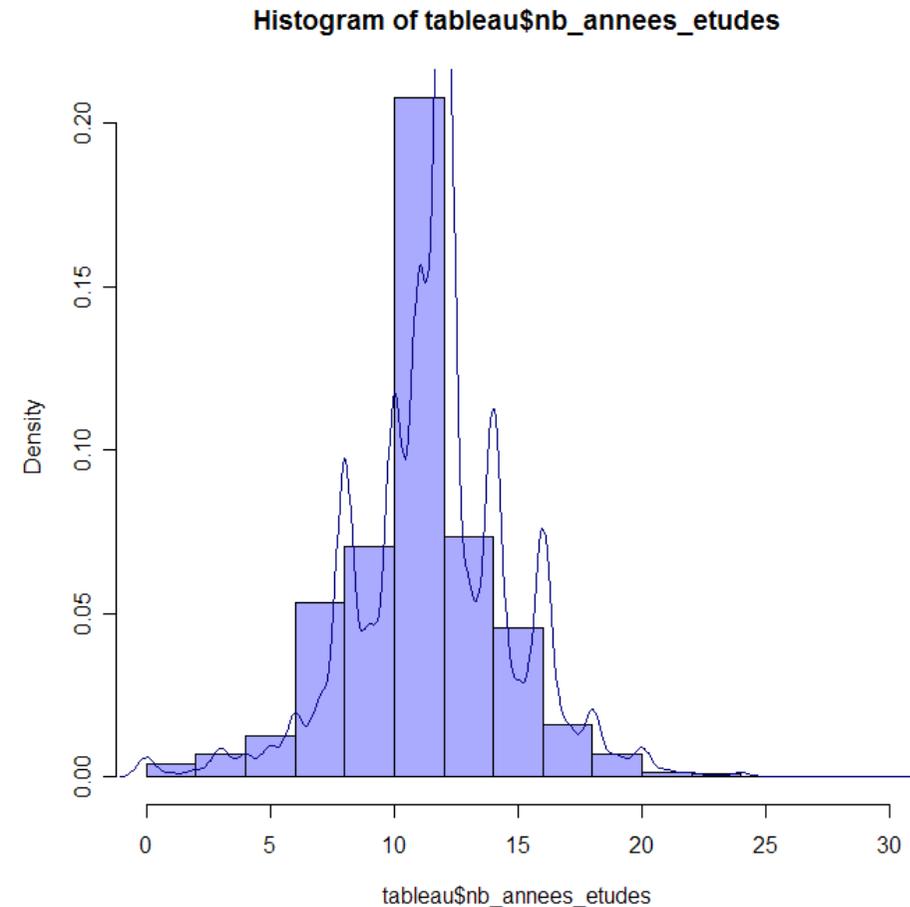
```
> plot(tableau$nb_annees_etudes) ;
```



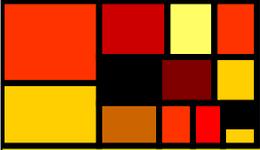
Variable quantitative discrète : Représentation de la distribution



```
> plot(tableau$nb_annees_etudes) ;
```

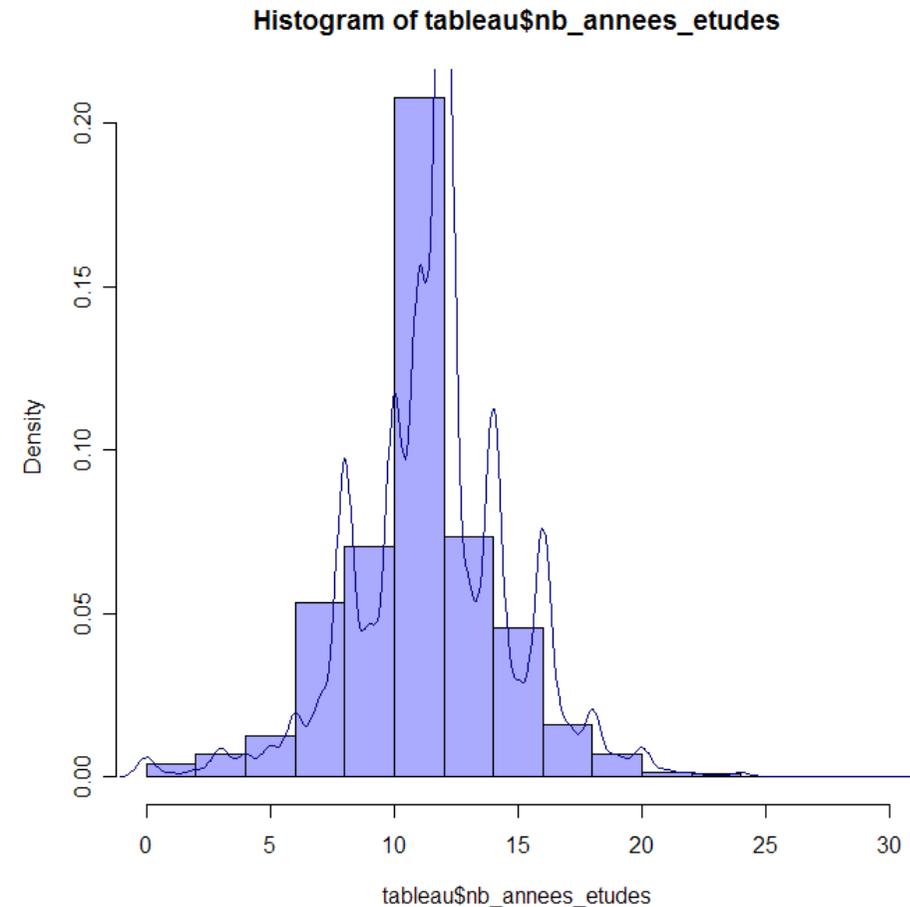


Variable quantitative discrète : Représentation de la distribution

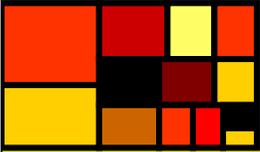


```
> plot(tableau$nb_annees_etudes) ;
```

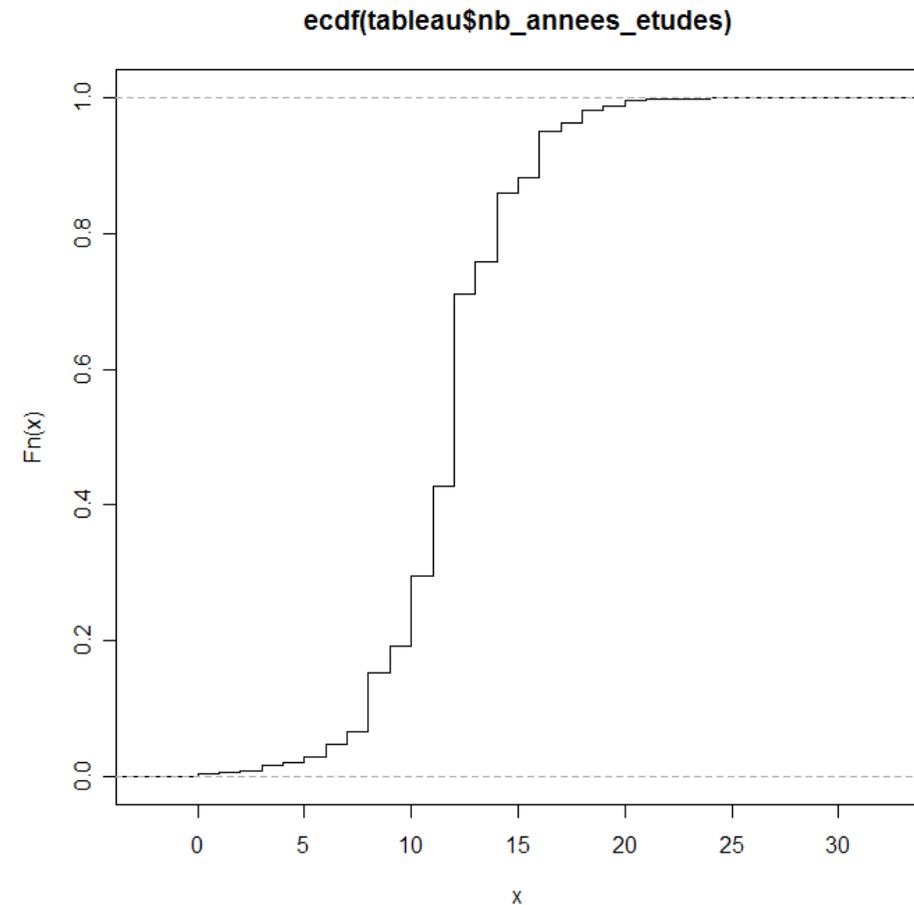
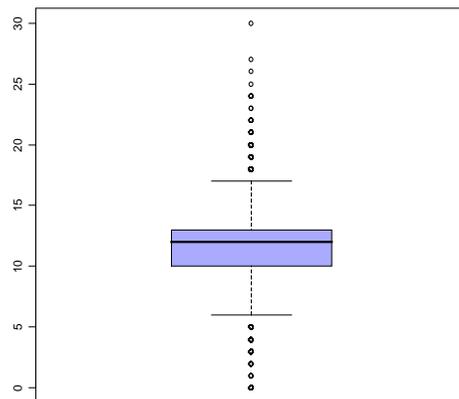
Dans le cas discret, pour un nombre de modalités élevé, l'histogramme est souvent accepté. Pourtant comme le montre l'estimation de la densité de probabilité, cette représentation est très critiquable.



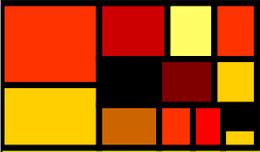
Variable quantitative discrète : Représentation de la distribution



- > `boxplot(tableau$nb_annees_etudes, col="#AAAAFF") ;`
- > `plot(ecdf(tableau$nb_annees_etudes), verticals=TRUE, do.points=FALSE) ;`

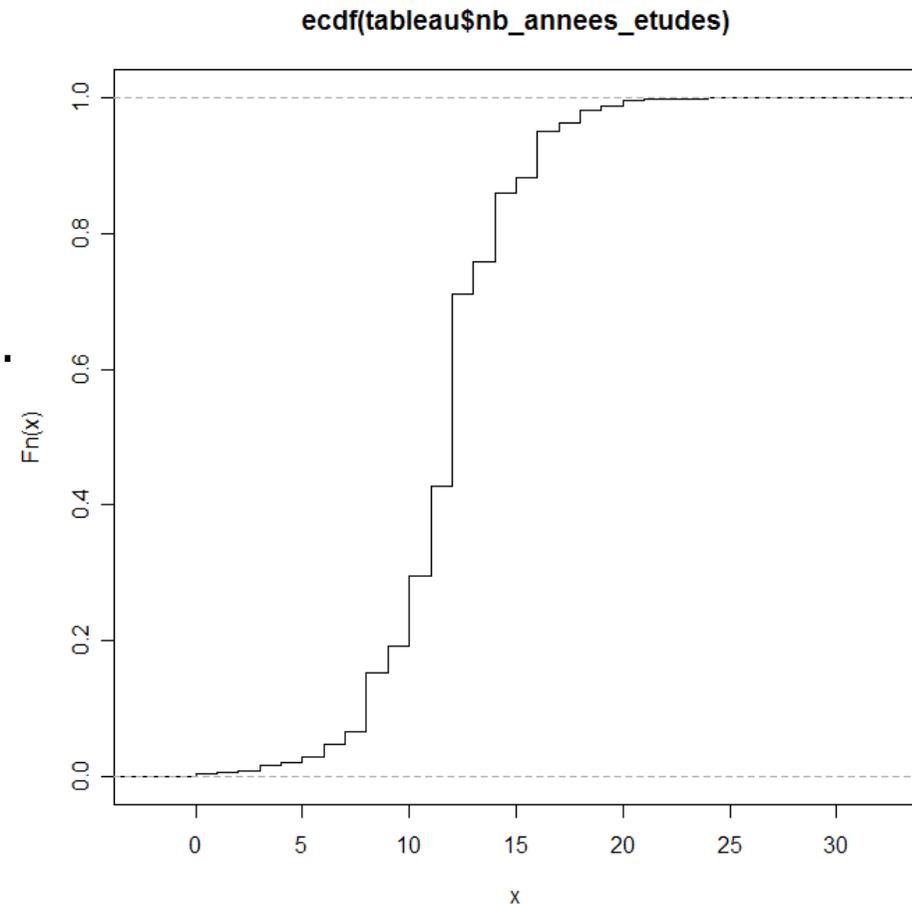
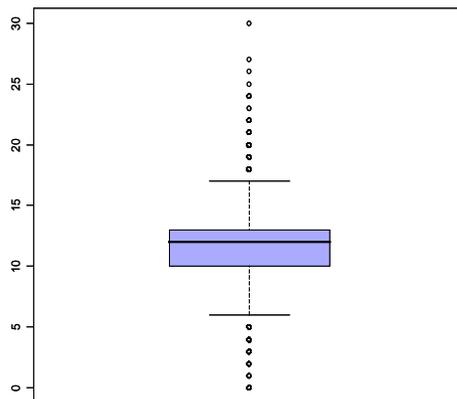


Variable quantitative discrète : Représentation de la distribution

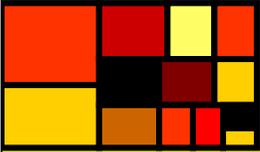


- > `boxplot(tableau$nb_annees_etudes, col="#AAAAFF") ;`
- > `plot(ecdf(tableau$nb_annees_etudes), verticals=TRUE, do.points=FALSE) ;`

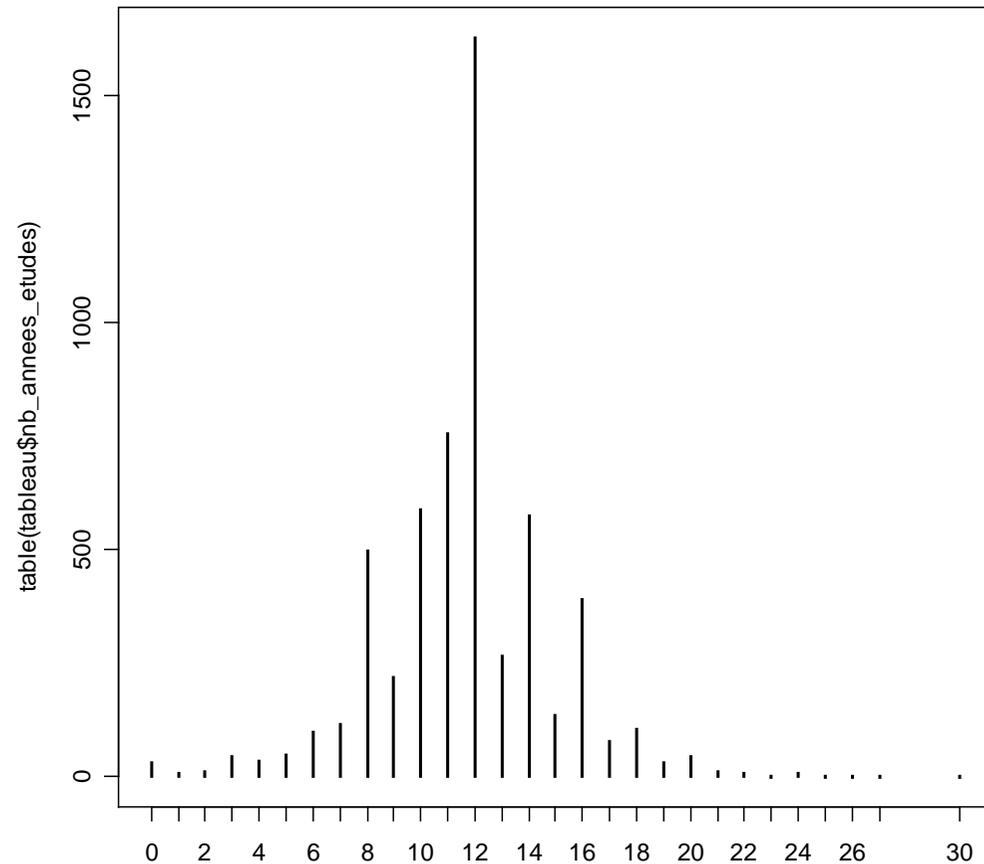
Boxplots et ECDF restent entièrement licites.



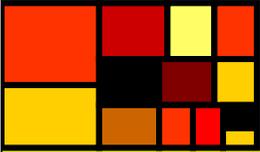
Variable quantitative discrète : Représentation de la distribution



```
> plot(table(tableau$nb_annees_etudes)) ;
```



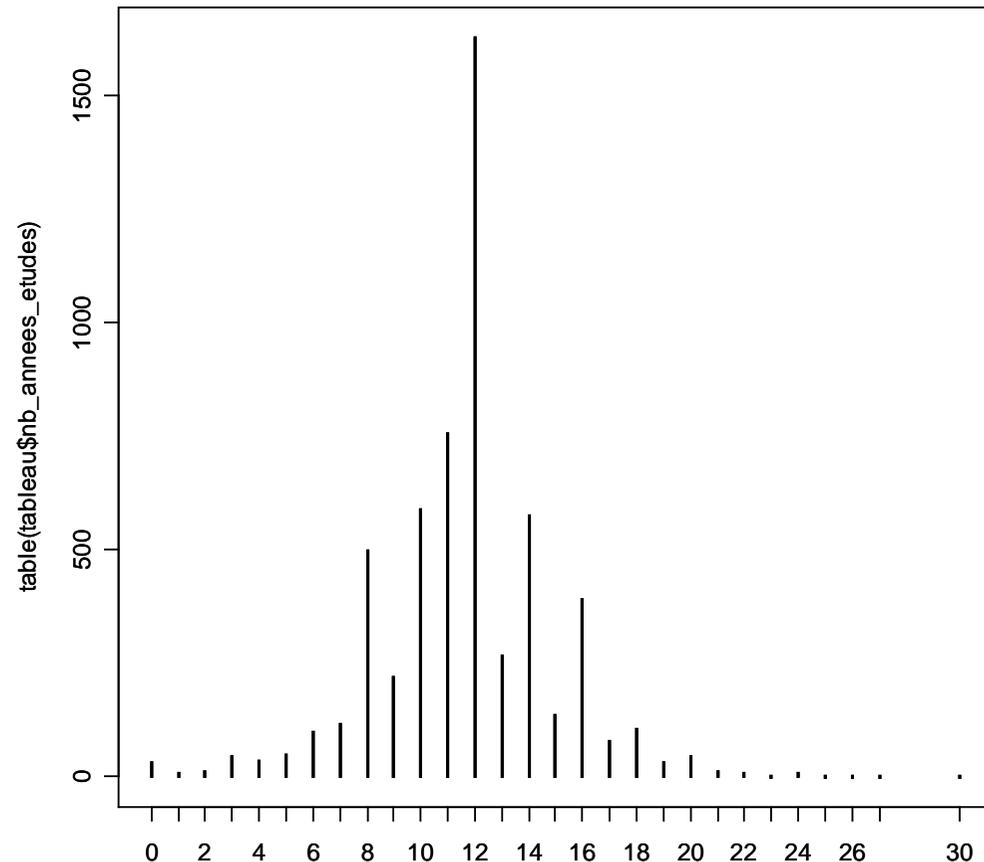
Variable quantitative discrète : Représentation de la distribution



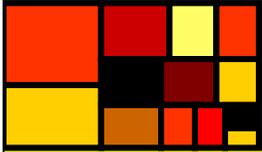
```
> plot(table(tableau$nb_annees_etudes)) ;
```

Le diagramme en bâtons du tableau de contingence est une option sans risque, même si l'aspect se complexifie avec un nombre de modalités important.

Remarque : l'abscisse demeure une véritable variable quantitative, ce qu'Excel ne sait pas faire.



Variable qualitative : représentation native

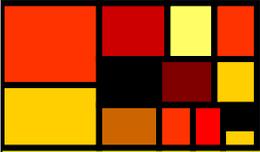


- Il n'existe pas de représentation native des variables qualitatives.
- Intuitivement et sans comprendre la différence, tous les utilisateurs représentent le tableau de contingence.

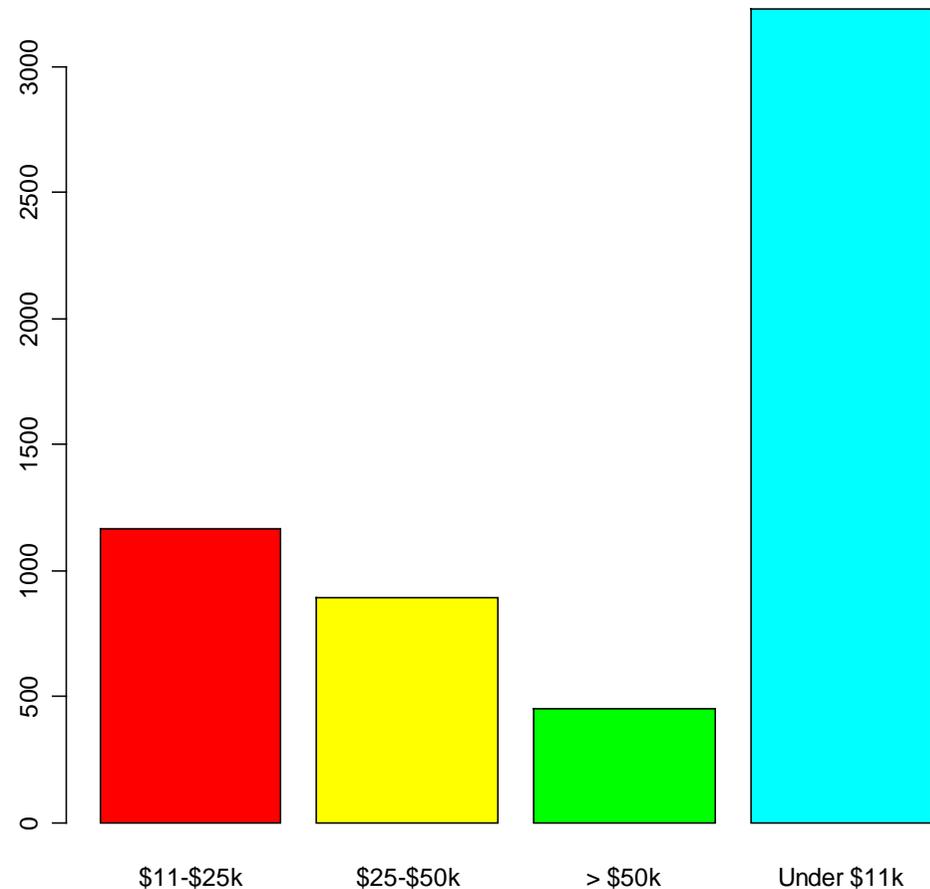
assurance	Nombre
Medicaid	647
Medicare	1458
Medicare & Medicaid	374
No insurance	322
Private	1698
Private & Medicare	1236

Variable qualitative :

Représentation de la distribution par diagramme en barres

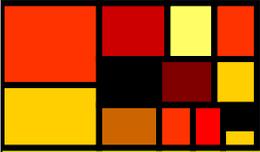


```
> barplot(table(tableau$revenus), col=rainbow(6)) ;
```



Variable qualitative :

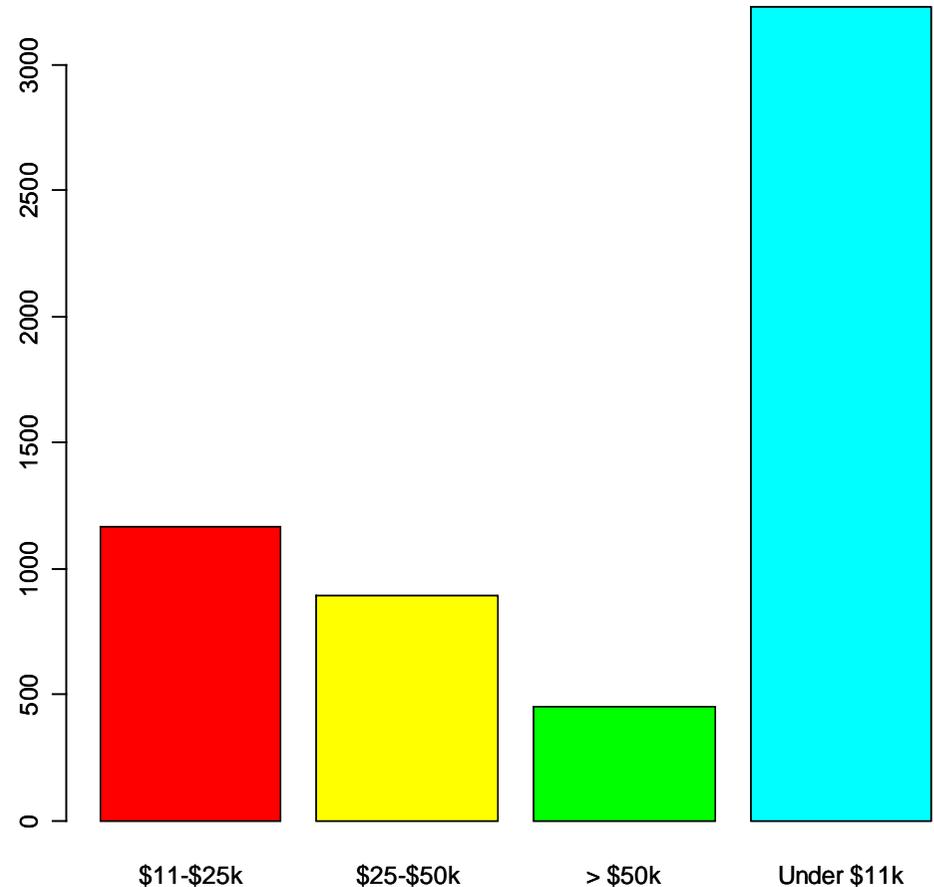
Représentation de la distribution par diagramme en barres



```
> barplot(table(tableau$revenus), col=rainbow(6)) ;
```

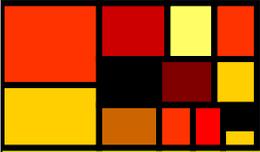
Abscisses = nom des classes
Ordonnées = effectif

Représentation « sans risque »,
mais pour un nombre de classes
élevé, elle fait perdre la notion de
proportion en faisant disparaître
la notion de « partition » (i.e.
somme=100%).

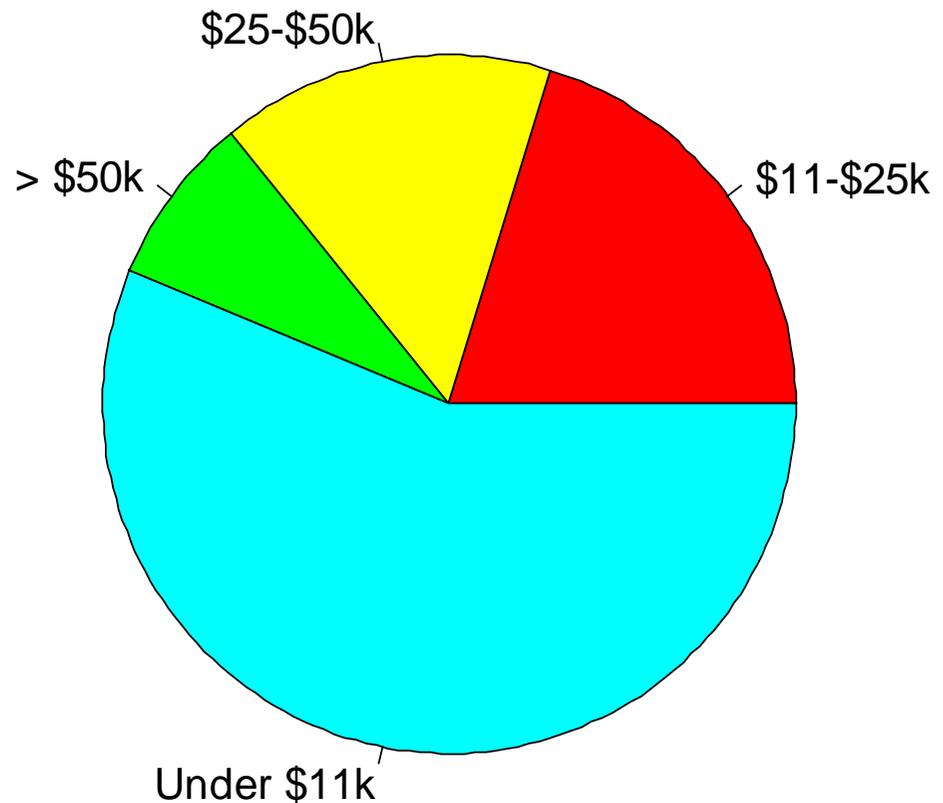


Variable qualitative :

Représentation de la distribution par diagramme en secteurs

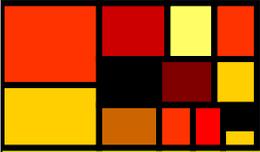


```
> pie(table(tableau$revenus),  
col=rainbow(6), cex=1.5)
```



Variable qualitative :

Représentation de la distribution par diagramme en secteurs



```
> pie(table(tableau$revenus),  
      col=rainbow(6), cex=1.5)
```

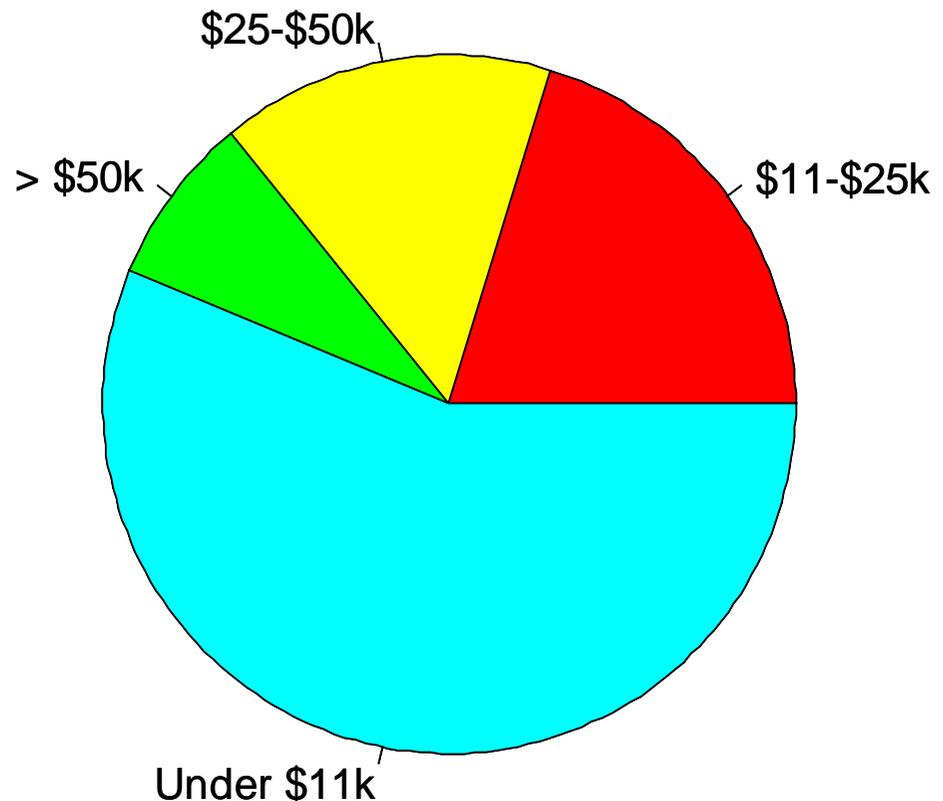
« camembert », très populaire dans le monde administratif. Effectif représenté par l'angle au centre (et donc la surface).

Avantages :

- explicitement « partition »
- Mise en évidence des grands ensembles et des proportions

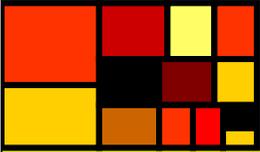
Inconvénients :

- perte de lisibilité pour les petites quantités
- Étiquetage indispensable mais parfois difficile
- Impossibilité de forcer la représentation des effectifs nuls.



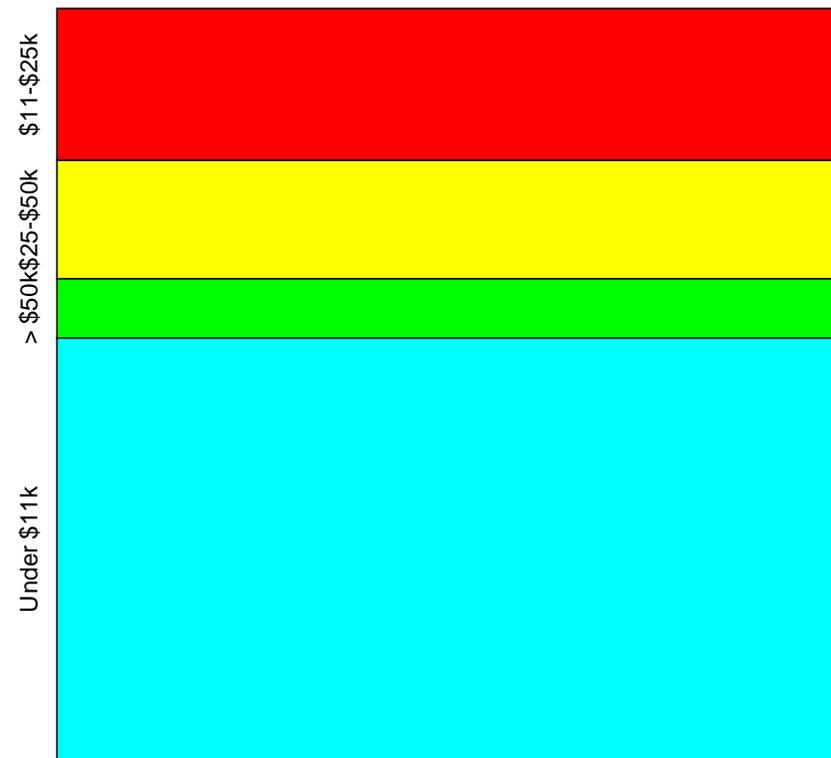
Variable qualitative :

Représentation de la distribution par barres empilées

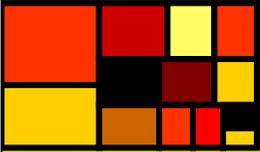


```
> mosaicplot(table(tableau$revenus), dir="h",  
color=rainbow(6), off=0, cex=1)
```

table(tableau\$revenus)



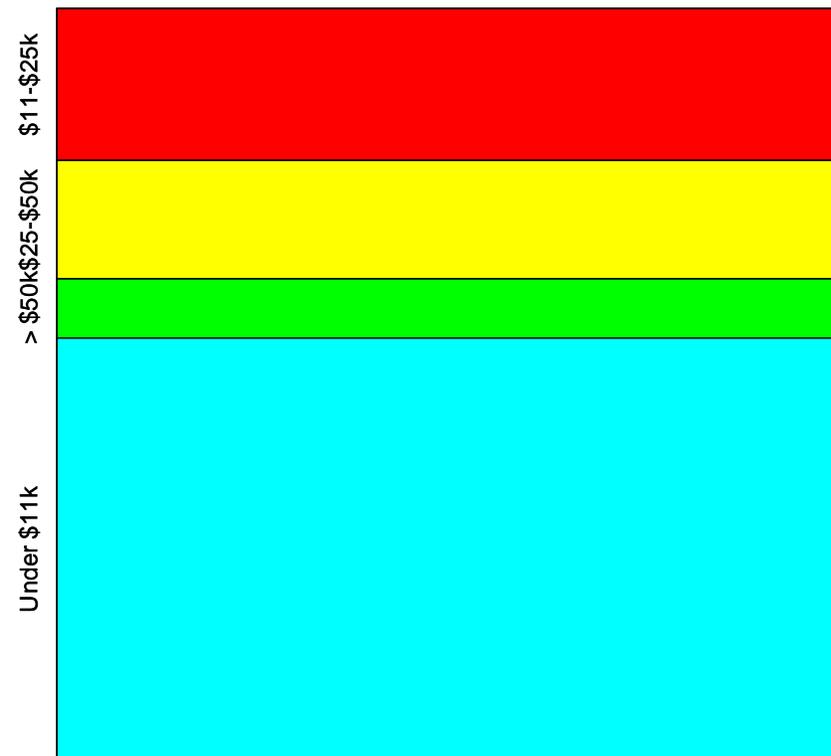
Variable qualitative : Représentation de la distribution par barres empilées



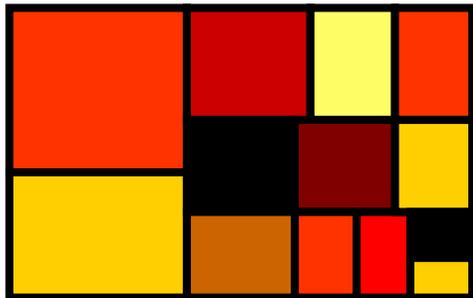
table(tableau\$revenus)

```
> mosaicplot(table(tableau$revenus), dir="h",  
color=rainbow(6), off=0, cex=1)
```

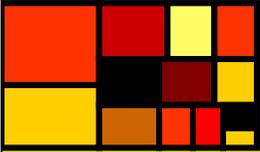
Peu esthétique sur un tableau de contingence à 1 dimension, la barre empilée n'est pas à négliger. Elle fait bien apparaître la notion de partition. Elle produit des proportions faciles à évaluer et à additionner.



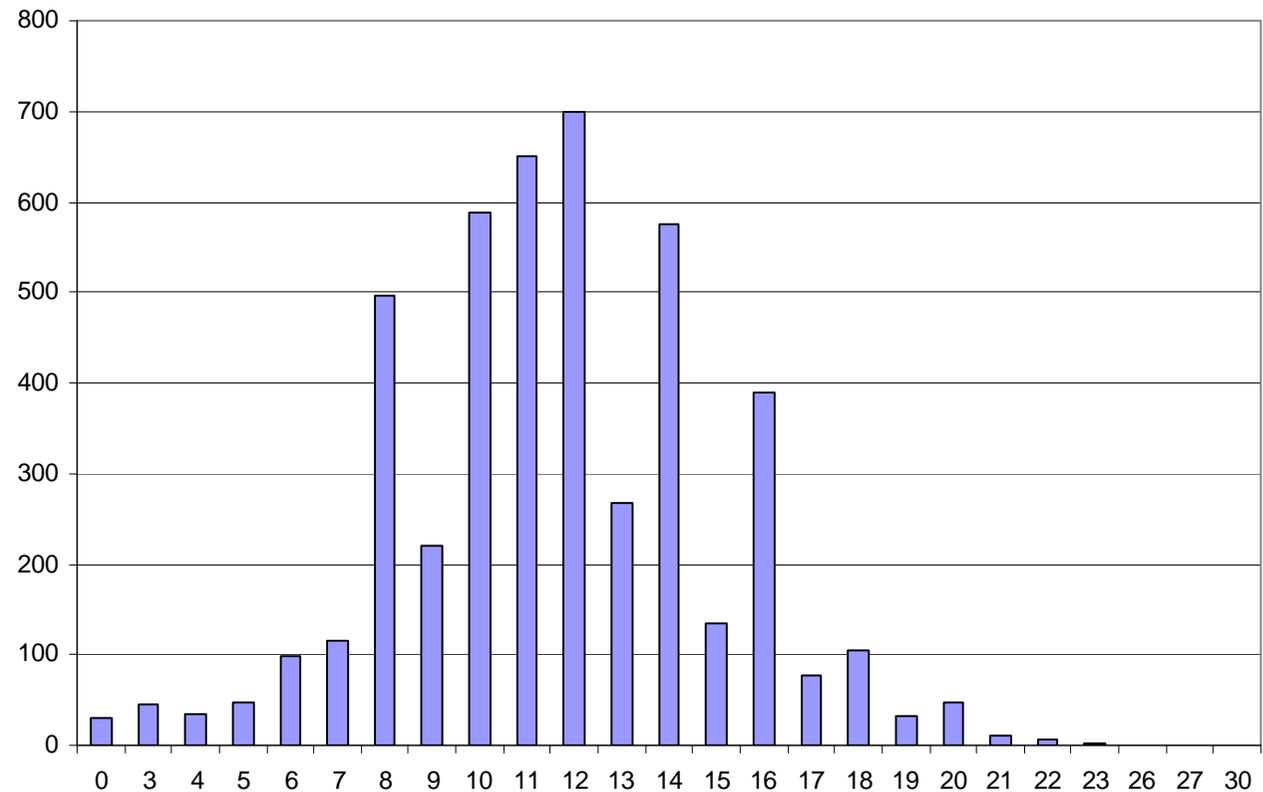
Représentations univariées : le petit musée des horreurs



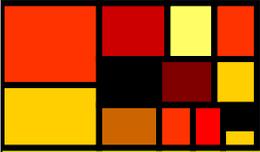
Le musée des horreurs, Variable quantitative discrète



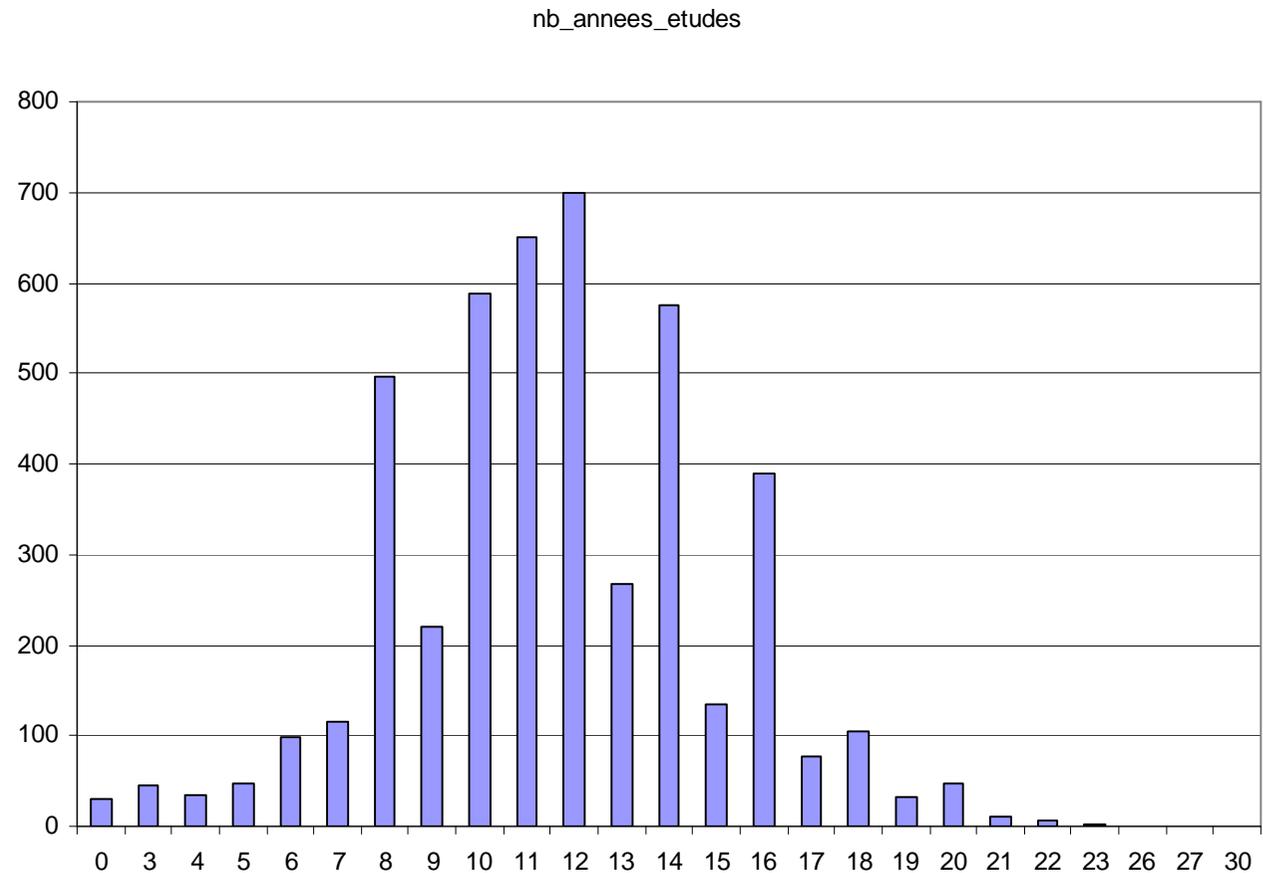
nb_annees_etudes



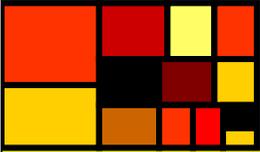
Le musée des horreurs, Variable quantitative discrète



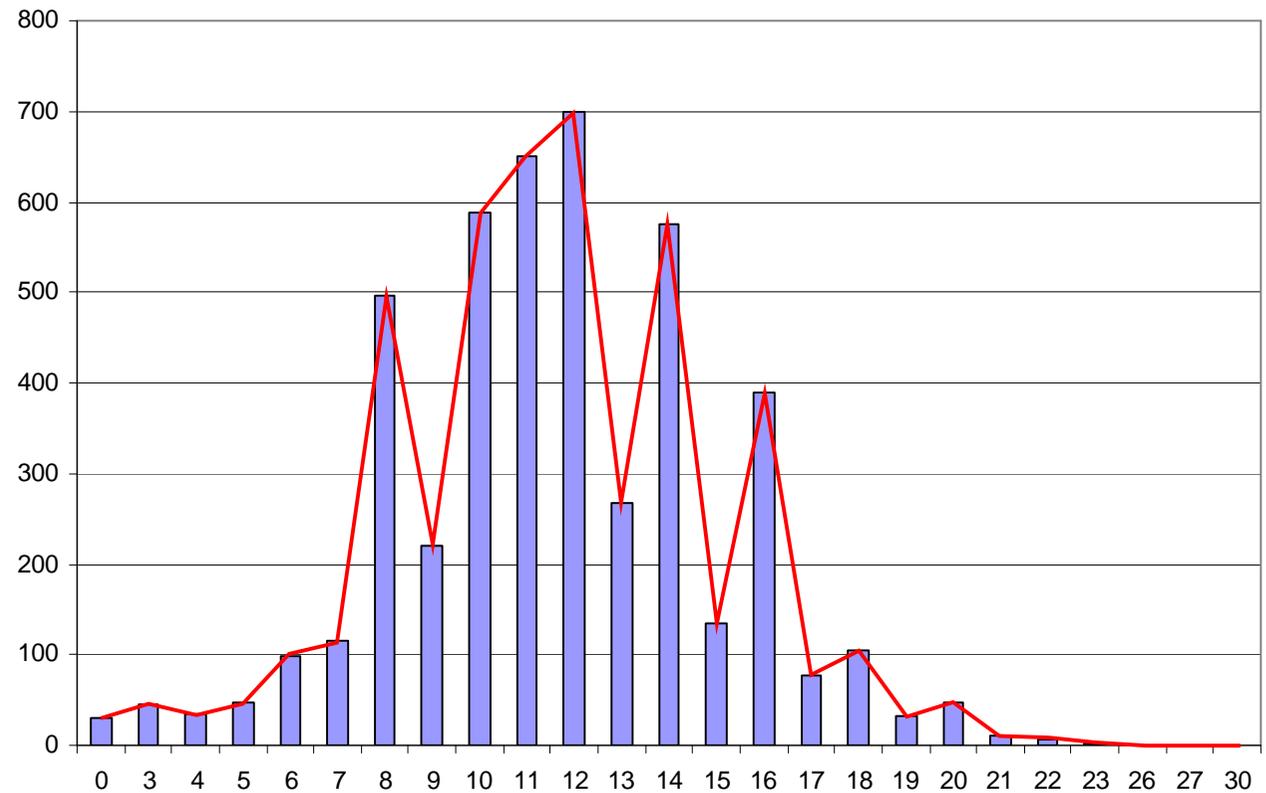
Ce diagramme en bâtons traite la variable comme une variable qualitative : les valeurs 0 et 3, 27 et 30 se suivent sans état d'âme.



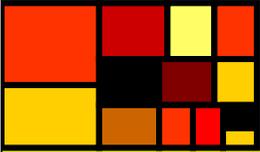
Le musée des horreurs, Variable quantitative discrète



nb_annees_etudes

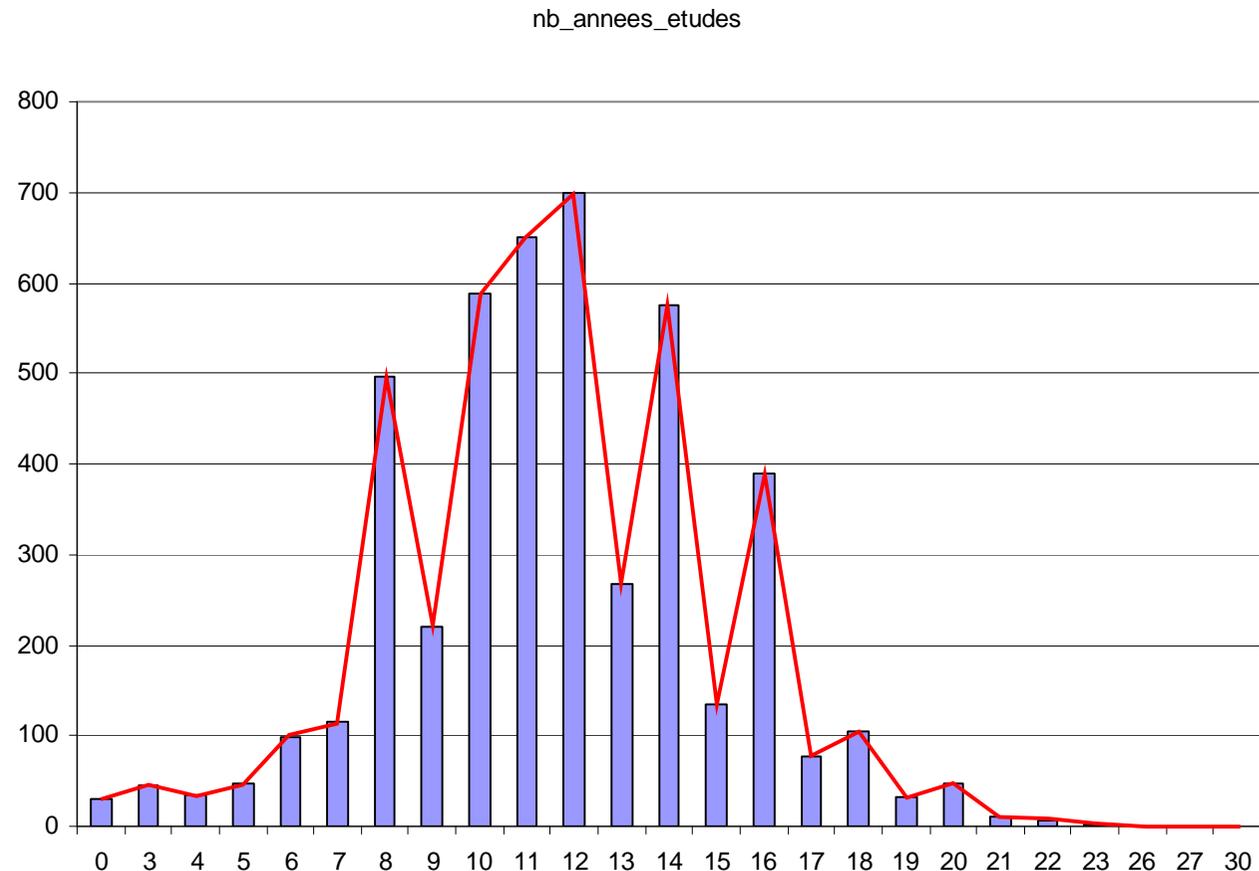


Le musée des horreurs, Variable quantitative discrète

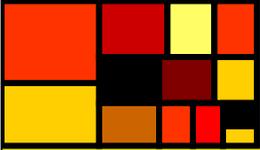


Ici en plus, les lignes suggèrent qu'une interpolation linéaire est licite. Ainsi on aurait le droit d'extrapoler le nombre d'individus ayant la valeur 9.1 ou la valeur 9.2... Ainsi le nombre d'individus est infini.

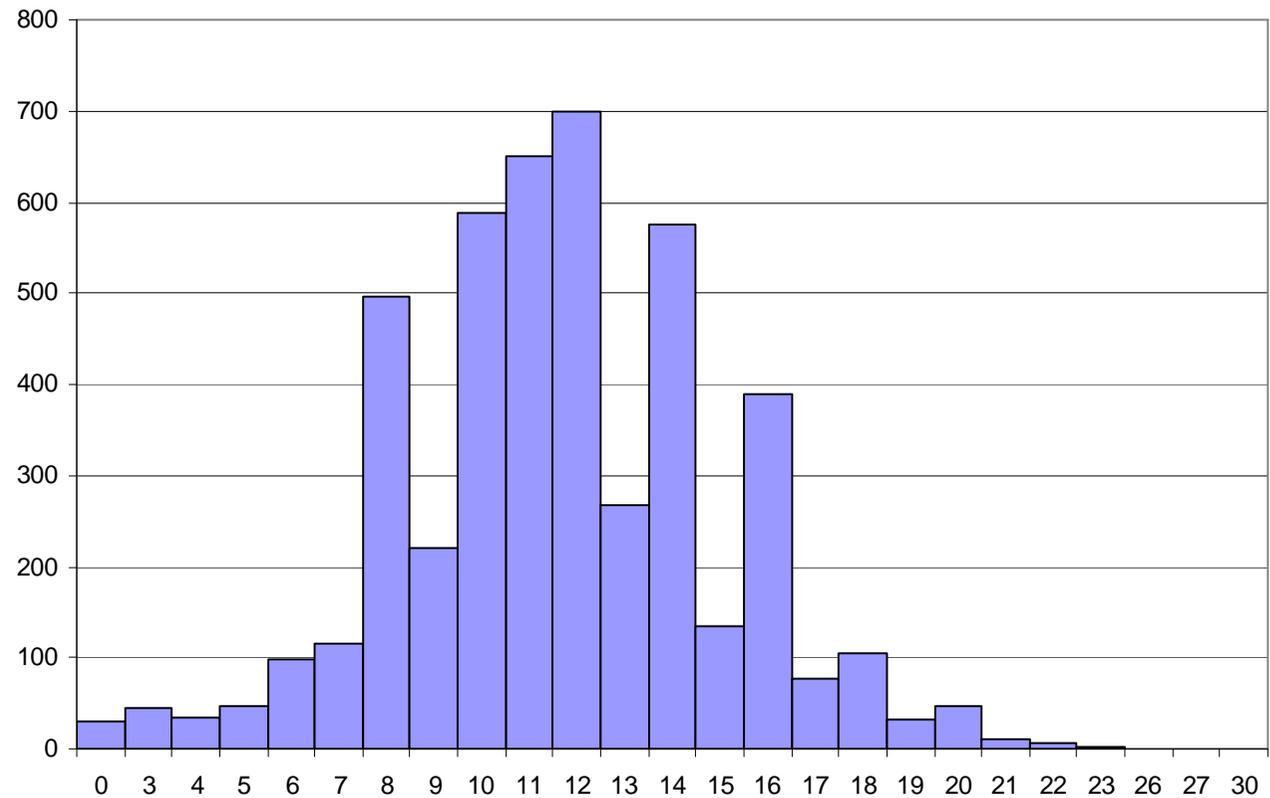
Relier les points n'est licite que dans certains cas bivariés, lorsqu'une interpolation est licite. Exemple : suivi de l'âge moyen mois après mois.



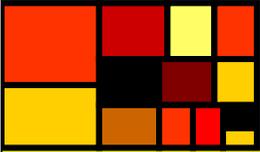
Le musée des horreurs, Variable quantitative discrète



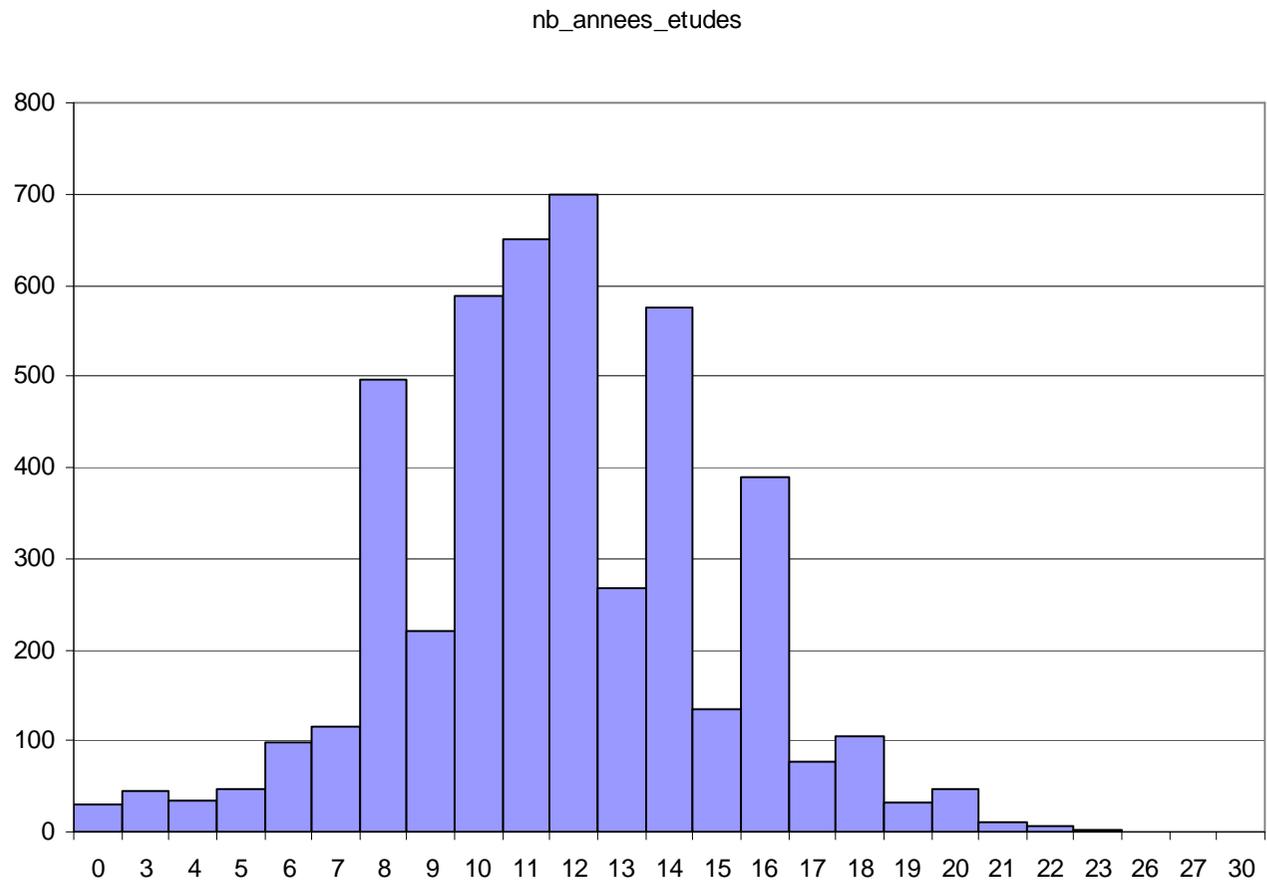
nb_annees_etudes



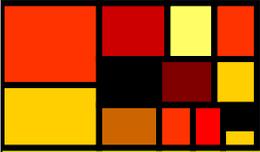
Le musée des horreurs, Variable quantitative discrète



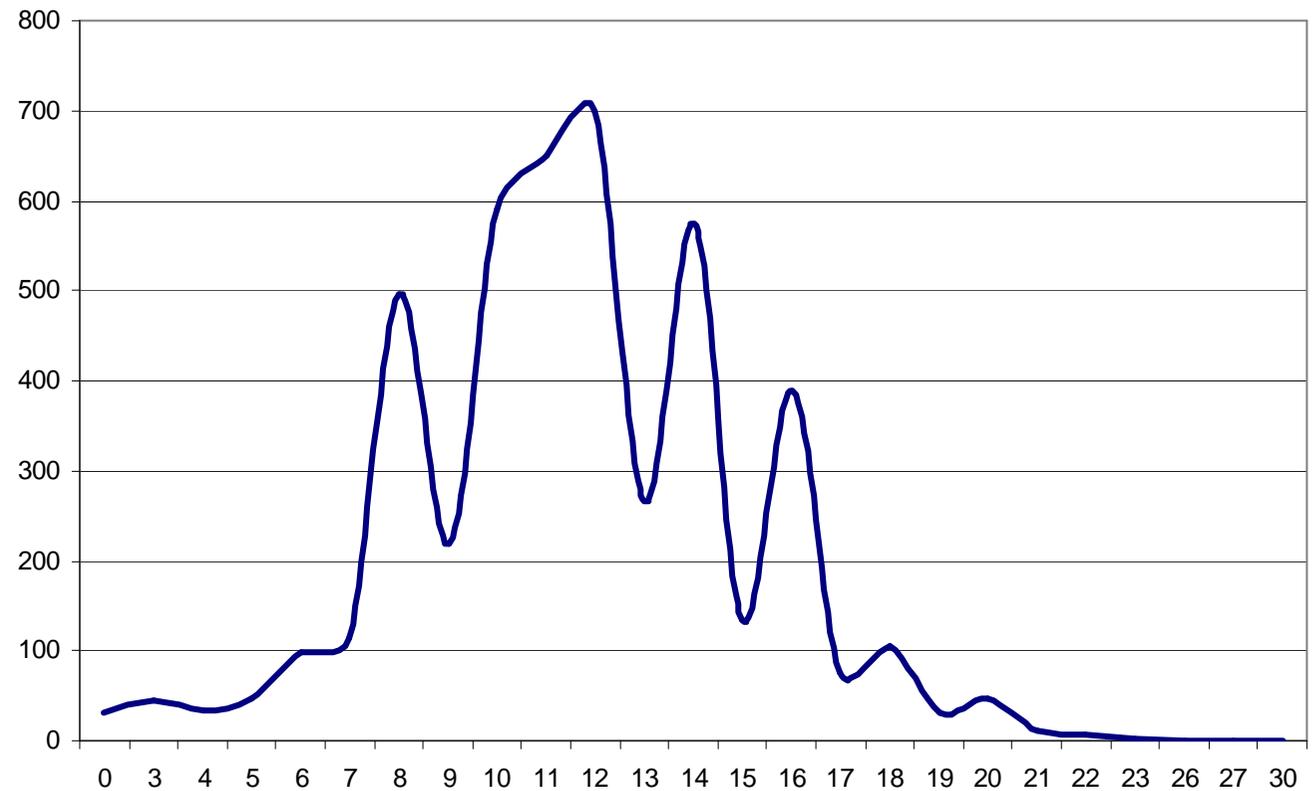
Pas soucis esthétique, les barres ont été élargies et se touchent. Ce procédé est exclusivement réservé aux histogrammes, or ceci est un diagramme en bâtons.



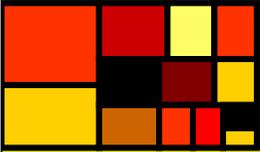
Le musée des horreurs, Variable quantitative discrète



nb_annees_etudes



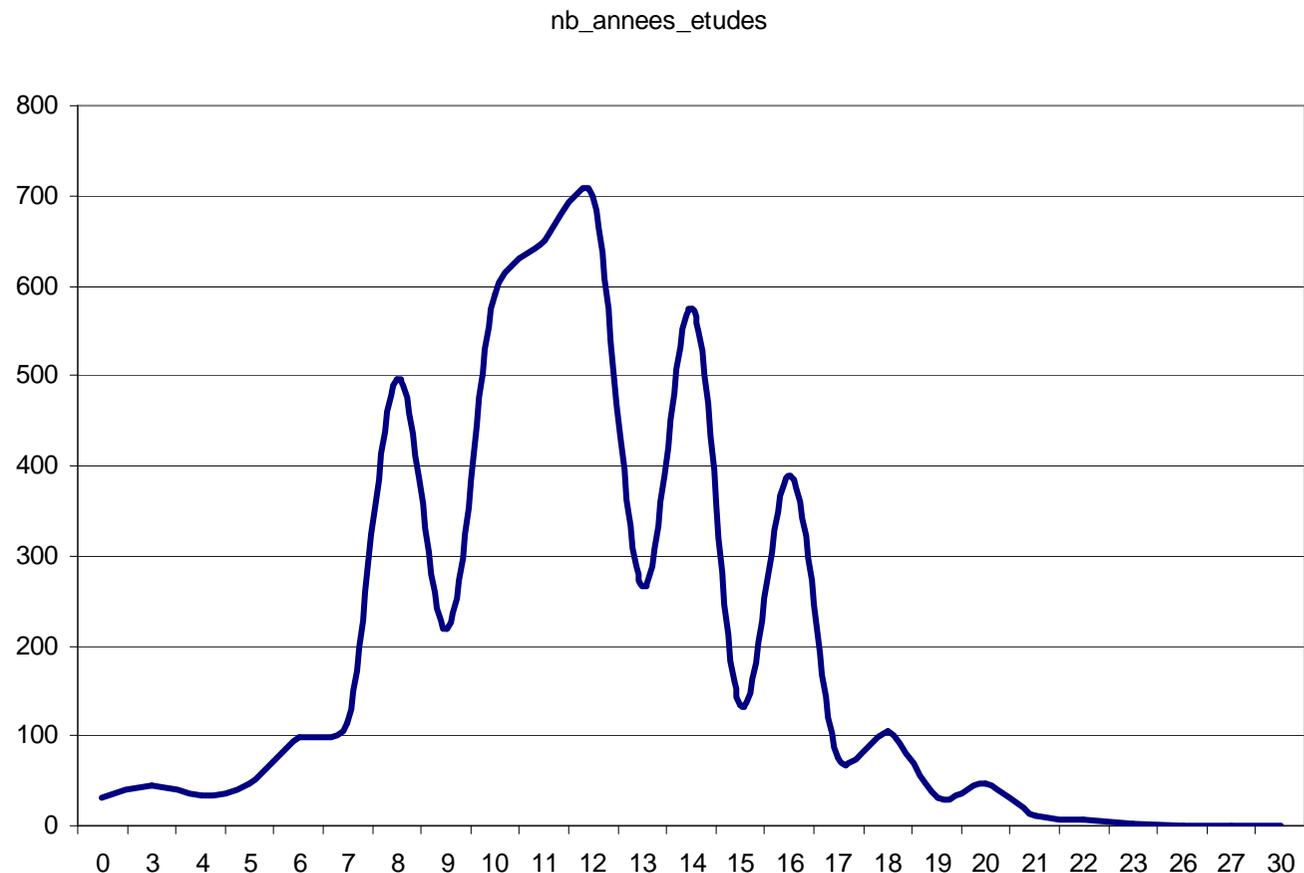
Le musée des horreurs, Variable quantitative discrète



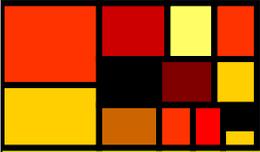
De pire en pire mais véridique, le dessin d'une courbe suggère une densité de probabilité.

Or :

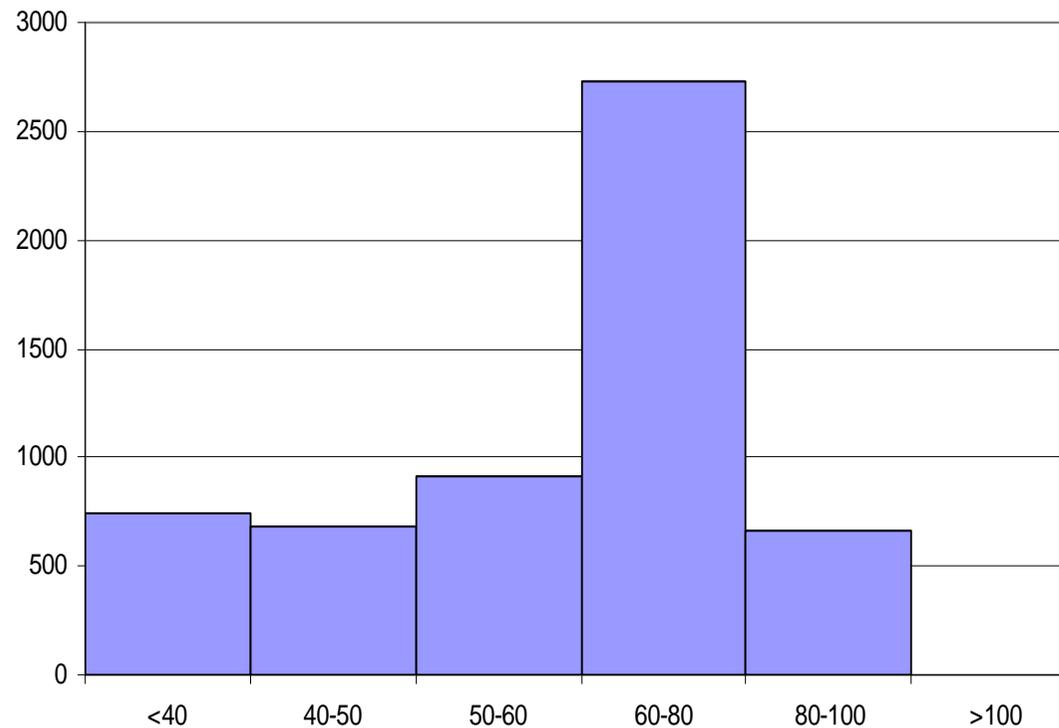
- l'abscisse n'est toujours pas quantitative
- la variable étant discrète, la densité de probabilité devrait être nulle entre les valeurs entières
- la surface sous la courbe est ici nettement supérieure à 1, ce qui est impossible.



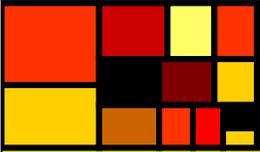
Le musée des horreurs, Variable quantitative continue



Répartition de l'âge



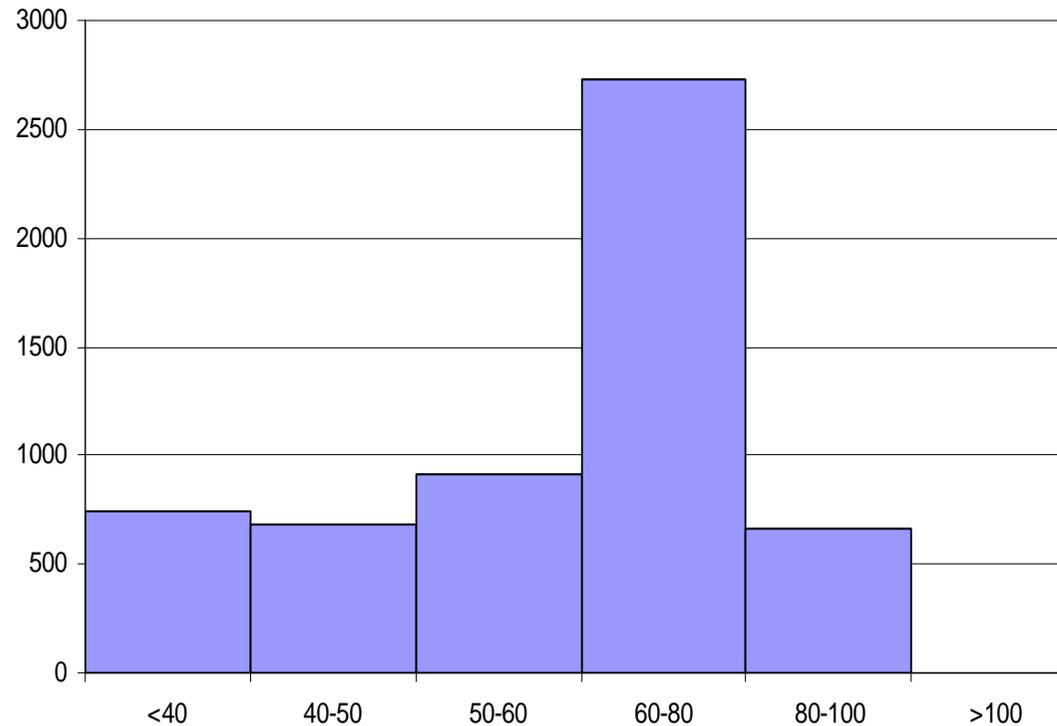
Le musée des horreurs, Variable quantitative continue



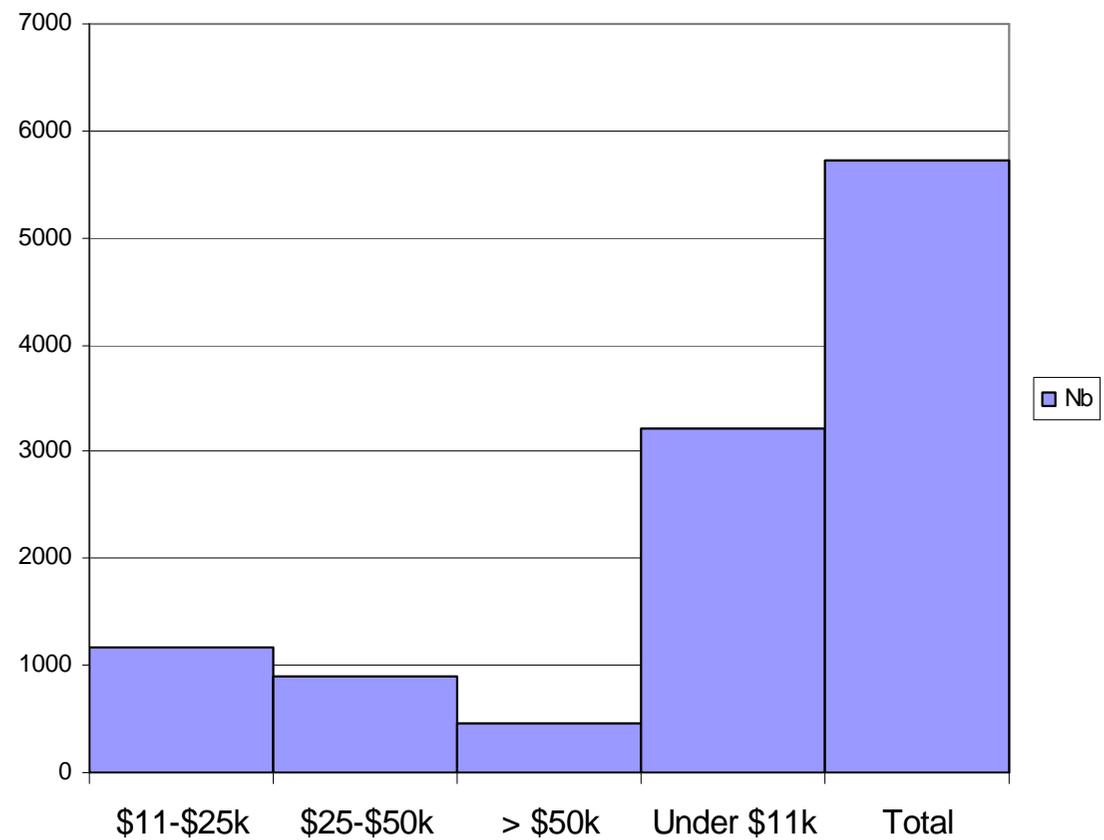
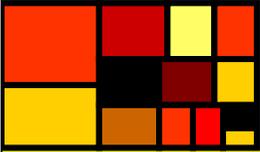
Histogramme faux :

- l'axe des abscisses n'est pas quantitatif (largeurs de classes variable)
- en cas de classes inégales, il est impératif d'exprimer l'ordonnée en proportion (effectifs tolérés seulement si classe égales). C'est alors la surface des rectangles qui importe et non leur hauteur.

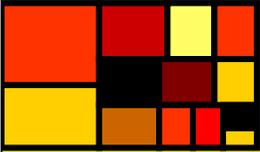
Répartition de l'âge



Le musée des horreurs, Variable qualitative



Le musée des horreurs, Variable qualitative



Fautes :

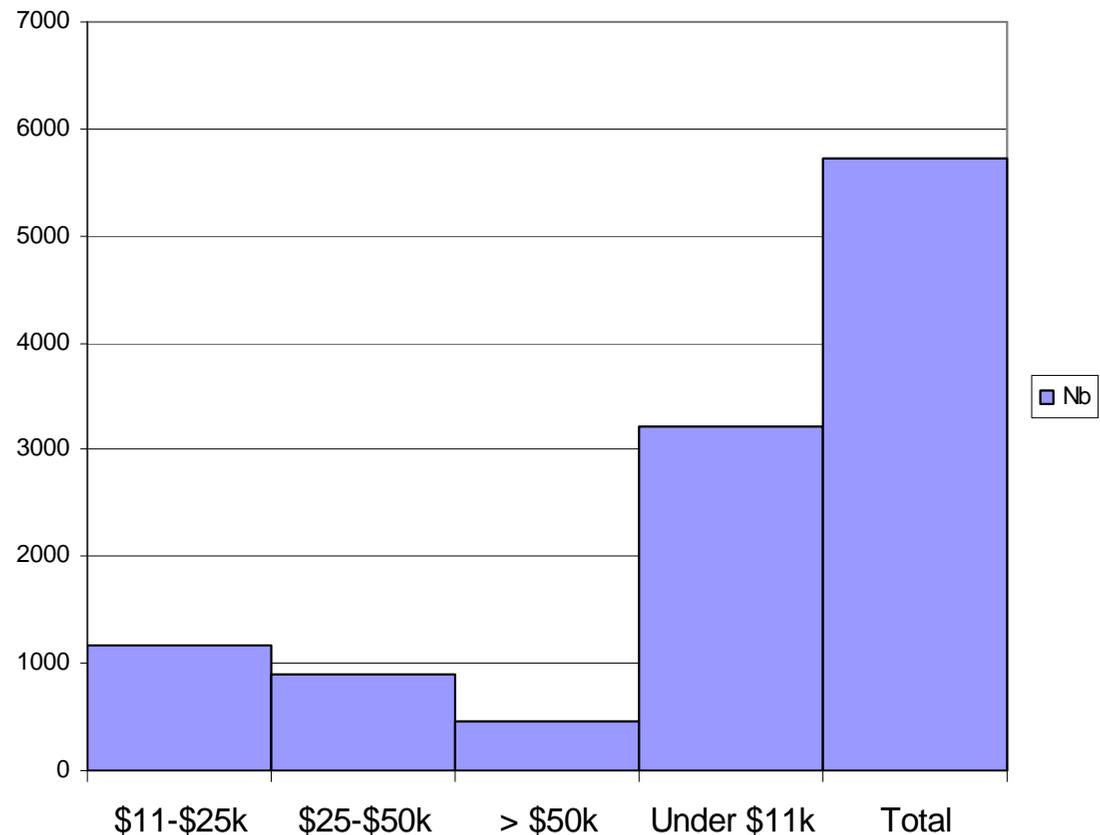
- les barres sont accolées alors que ce n'est pas un histogramme

- les modalités sont triées alphabétiquement alors qu'il s'agit d'une variable qualitative ordonnée

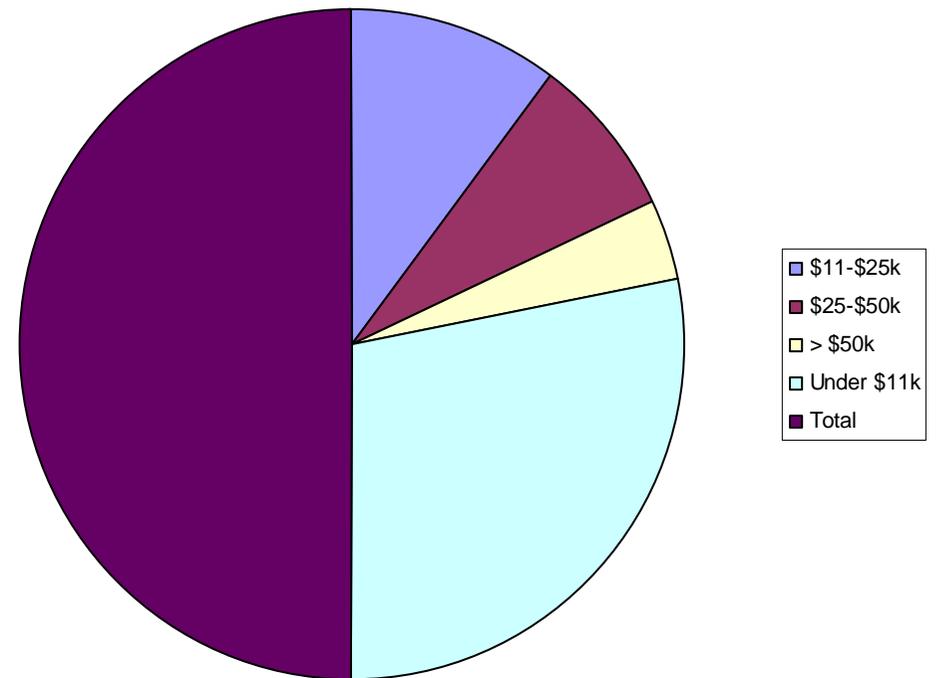
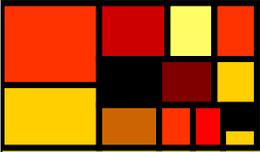
- le total apparaît sur le même graphe

Imperfection :

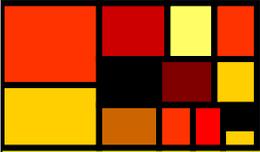
- ce graphique ne fait pas comprendre la notion de partition



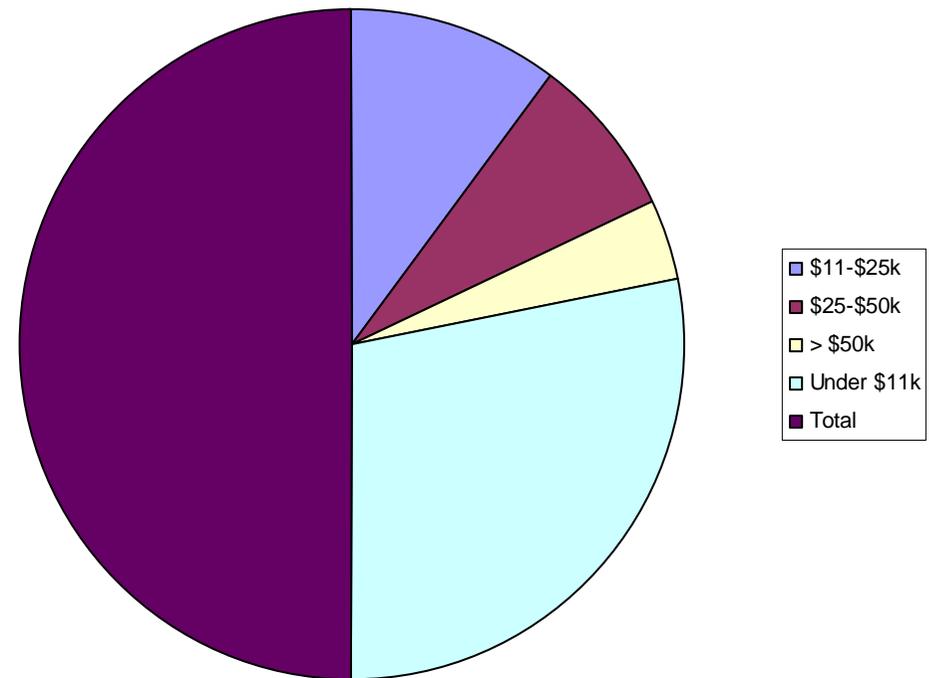
Le musée des horreurs, Variable qualitative



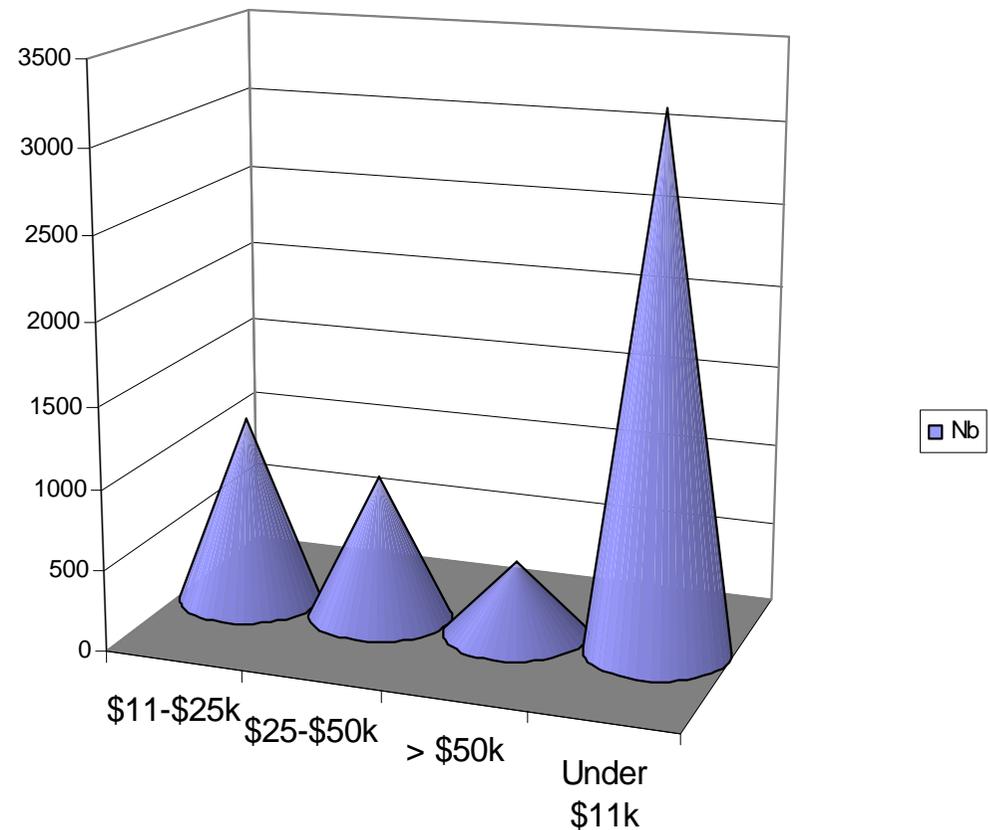
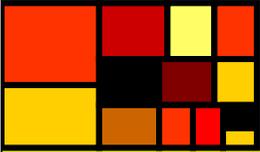
Le musée des horreurs, Variable qualitative



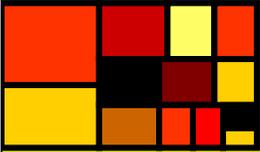
Archi-faux, le total apparaît en même temps que les modalités, sur un graphe pourtant spécialement dédié aux partitions.



Le musée des horreurs, Variable qualitative

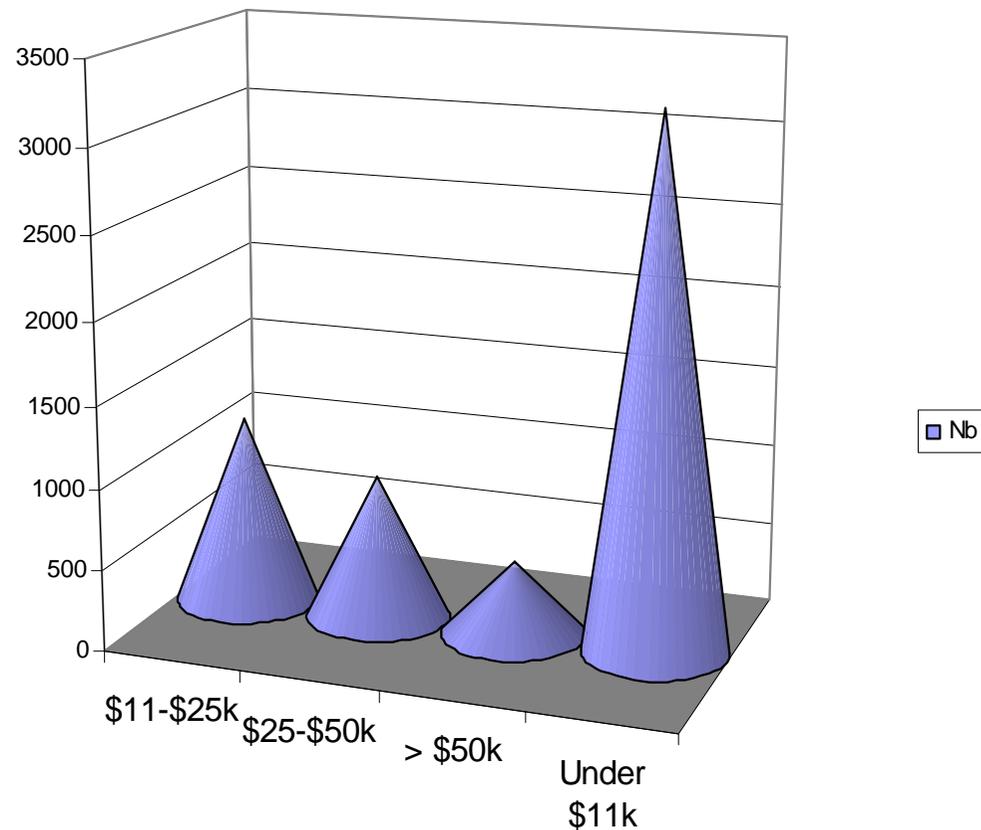


Le musée des horreurs, Variable qualitative



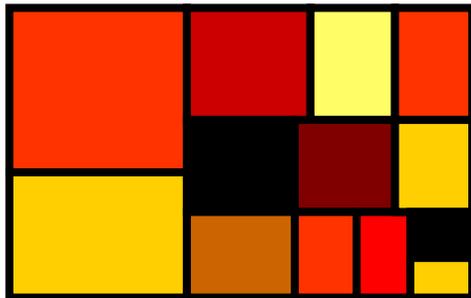
Fautes :

- problème de classement des modalités
- effet 3D douteux qui ne permet plus de lire les quantités
- la quantité est représentée par la hauteur... mais nullement par la surface ou le volume suggéré des pictogrammes !



Préalable aux représentations multivariées :

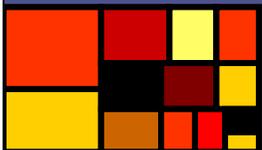
précisions sur les notions de distribution et d'agrégation



- I. Cas univarié
- II. Cas bivarié (et multivarié)

Cas univarié

variable quantitative discrète :
données natives, distribution , agrégation

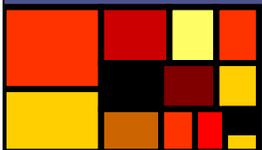


nb_annees_etudes
12
12
14
9
10
8
14
12
13
11
15
12
12
16
3
8
14
...

Données natives

Cas univarié

variable quantitative discrète : données natives, distribution , agrégation



nb_annees_etudes
12
12
14
9
10
8
14
12
13
11
15
12
12
16
3
8
14
...

Données natives

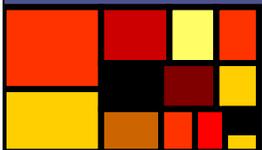
nb_annees_e	Effectif
0	31
3	45
4	34
5	47
6	98
7	115
8	497
9	220
10	589
11	650
12	700
13	267
14	575
15	134
16	389
17	77
18	105
19	32
20	46
21	11
22	7
23	2
26	1
27	1
30	1

Distribution (tableau de contingence)

Cas univarié

variable quantitative discrète :

données natives, distribution , agrégation



nb_annees_etudes
12
12
14
9
10
8
14
12
13
11
15
12
12
16
3
8
14
...

Données natives

nb_annees_e	Effectif
0	31
3	45
4	34
5	47
6	98
7	115
8	497
9	220
10	589
11	650
12	700
13	267
14	575
15	134
16	389
17	77
18	105
19	32
20	46
21	11
22	7
23	2
26	1
27	1
30	1

Distribution (tableau de contingence)

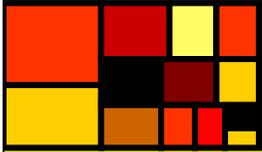
moyenne	12
médiane	12
minimum	0
maximum	30
mode	12
1° quartile	10
3° quartile	13
95° percentile	17
variance	9.982228278
écart type	3.159466455
décompte*	5735

Exemples d'agrégations

* Le décompte est une forme d'agrégation en SQL, mais par soucis didactique et puisqu'il n'utilise pas la variable elle-même, nous le présentons sous le concept de « distribution »

Cas univarié

variable quantitative continue :
données natives, distribution , agrégation

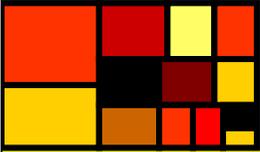


age
70.2509766
78.1789551
46.09198
75.3319702
67.9099731
86.0779419
54.9679871
43.6389771
18.0419922
48.4239807
34.4419861
68.3479614
74.7099609
42.2369995
81.9709473
78.3049927
88.4219971
...

Données natives

Cas univarié

variable quantitative continue : données natives, distribution , agrégation



age
70.2509766
78.1789551
46.09198
75.3319702
67.9099731
86.0779419
54.9679871
43.6389771
18.0419922
48.4239807
34.4419861
68.3479614
74.7099609
42.2369995
81.9709473
78.3049927
88.4219971
...

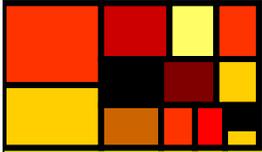
Il n'est pas possible de représenter la distribution d'une variable continue par tableau. On ne peut qu'utiliser :

- la densité de probabilité
- la fonction de répartition

Données natives

Cas univarié

variable quantitative continue : données natives, distribution , agrégation



age
70.2509766
78.1789551
46.09198
75.3319702
67.9099731
86.0779419
54.9679871
43.6389771
18.0419922
48.4239807
34.4419861
68.3479614
74.7099609
42.2369995
81.9709473
78.3049927
88.4219971
...

Données natives

Il n'est pas possible de représenter la distribution d'une variable continue par tableau. On ne peut qu'utiliser :
-la densité de probabilité
-la fonction de répartition

moyenne	61
médiane	64
minimum	18
maximum	102
mode	68
1° quartile	50.1549988
3° quartile	73.9324646
95° percentile	84.9975525
variance	278.217369
écart type	16.6798492
décompte*	5735

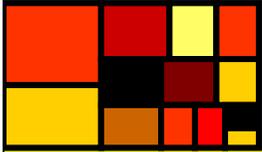
Exemples d'agrégations

* Le décompte est une forme d'agrégation en SQL, mais par soucis didactique et puisqu'il n'utilise pas la variable elle-même, nous le présentons sous le concept de « distribution »

Cas univarié

variable qualitative :

données natives, distribution , agrégation



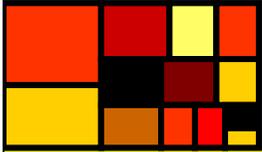
assurance
Medicare
Private & Medicare
Private
Private & Medicare
Medicare
Medicare
Private
Private
Private
Medicaid
Private
Private & Medicare
Private & Medicare
Private
Medicare
Medicare
Private & Medicare
...

Données natives

Cas univarié

variable qualitative :

données natives, distribution , agrégation



assurance
Medicare
Private & Medicare
Private
Private & Medicare
Medicare
Medicare
Private
Private
Private
Medicaid
Private
Private & Medicare
Private & Medicare
Private
Medicare
Medicare
Private & Medicare
...

assurance	Effectif
Medicaid	647
Medicare	1457
Medicare & Medicaid	374
No insurance	322
Private	1698
Private & Medicare	1236

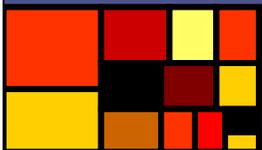
Données natives

Distribution
(tableau de contingence)

Cas univarié

variable qualitative :

données natives, distribution , agrégation



assurance
Medicare
Private & Medicare
Private
Private & Medicare
Medicare
Medicare
Private
Private
Private
Medicaid
Private
Private & Medicare
Private & Medicare
Private
Medicare
Medicare
Private & Medicare
...

Données natives

assurance	Effectif
Medicaid	647
Medicare	1457
Medicare & Medicaid	374
No insurance	322
Private	1698
Private & Medicare	1236

Distribution
(tableau de contingence)

Mode	Private
------	---------

Décompte*	5735
-----------	------

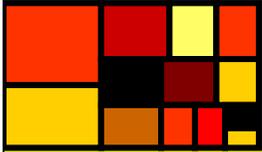
Seule agrégation possible : le mode

** Le décompte est une forme d'agrégation en SQL, mais par soucis didactique et puisqu'il n'utilise pas la variable elle-même, nous le présentons sous le concept de « distribution »*

Cas bivarié

qualitatif * qualitatif :

données natives, distribution , agrégation



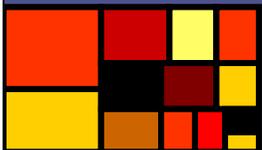
nb_annees_etudes
12
12
14
9
10
8
14
12
13
11
15
12
12
16
3
8
14
...

Données natives

Cas bivarié

qualitatif * qualitatif :

données natives, distribution , agrégation



nb_annees_etudes
12
12
14
9
10
8
14
12
13
11
15
12
12
16
3
8
14
...

revenus	assurance	effectif
Under \$11k	Medicaid	539
Under \$11k	Medicare	978
Under \$11k	Medicare & Medicaid	323
Under \$11k	No insurance	225
Under \$11k	Private	487
Under \$11k	Private & Medicare	674
\$11-\$25k	Medicaid	82
\$11-\$25k	Medicare	305
\$11-\$25k	Medicare & Medicaid	44
\$11-\$25k	No insurance	51
\$11-\$25k	Private	334
\$11-\$25k	Private & Medicare	348
\$25-\$50k	Medicaid	23
\$25-\$50k	Medicare	113
\$25-\$50k	Medicare & Medicaid	5
\$25-\$50k	No insurance	31
\$25-\$50k	Private	588
\$25-\$50k	Private & Medicare	133
> \$50k	Medicaid	3
> \$50k	Medicare	61
> \$50k	Medicare & Medicaid	2
> \$50k	No insurance	15
> \$50k	Private	289
> \$50k	Private & Medicare	81

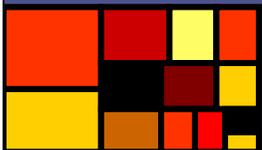
Données natives

Distribution jointe
(tableau de contingence)

Cas bivarié

qualitatif * qualitatif :

données natives, distribution , agrégation



nb_annees_etudes
12
12
14
9
10
8
14
12
13
11
15
12
12
16
3
8
14
...

Données natives

revenus	assurance	effectif
Under \$11k	Medicaid	539
Under \$11k	Medicare	978
Under \$11k	Medicare & Medicaid	323
Under \$11k	No insurance	225
Under \$11k	Private	487
Under \$11k	Private & Medicare	674
\$11-\$25k	Medicaid	82
\$11-\$25k	Medicare	305
\$11-\$25k	Medicare & Medicaid	44
\$11-\$25k	No insurance	51
\$11-\$25k	Private	334
\$11-\$25k	Private & Medicare	348
\$25-\$50k	Medicaid	23
\$25-\$50k	Medicare	113
\$25-\$50k	Medicare & Medicaid	5
\$25-\$50k	No insurance	31
\$25-\$50k	Private	588
\$25-\$50k	Private & Medicare	133
> \$50k	Medicaid	3
> \$50k	Medicare	61
> \$50k	Medicare & Medicaid	2
> \$50k	No insurance	15
> \$50k	Private	289
> \$50k	Private & Medicare	81

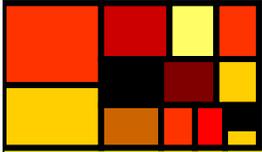
Distribution jointe
(tableau de contingence)

revenus	mode(assurance)
Under \$11k	Medicare
\$11-\$25k	Private & Medicare
\$25-\$50k	Private
> \$50k	Private

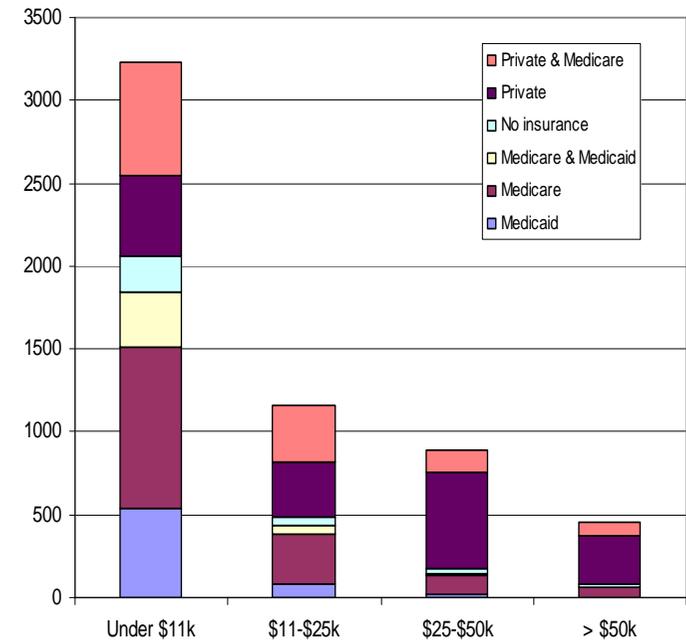
Exemple d'agrégations

Rappelons que, en dehors du décompte, seul le mode permet d'agréger une variable qualitative. On peut naturellement inverser les variables

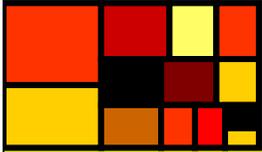
Cas bivarié distribution(qualitatif * qualitatif)



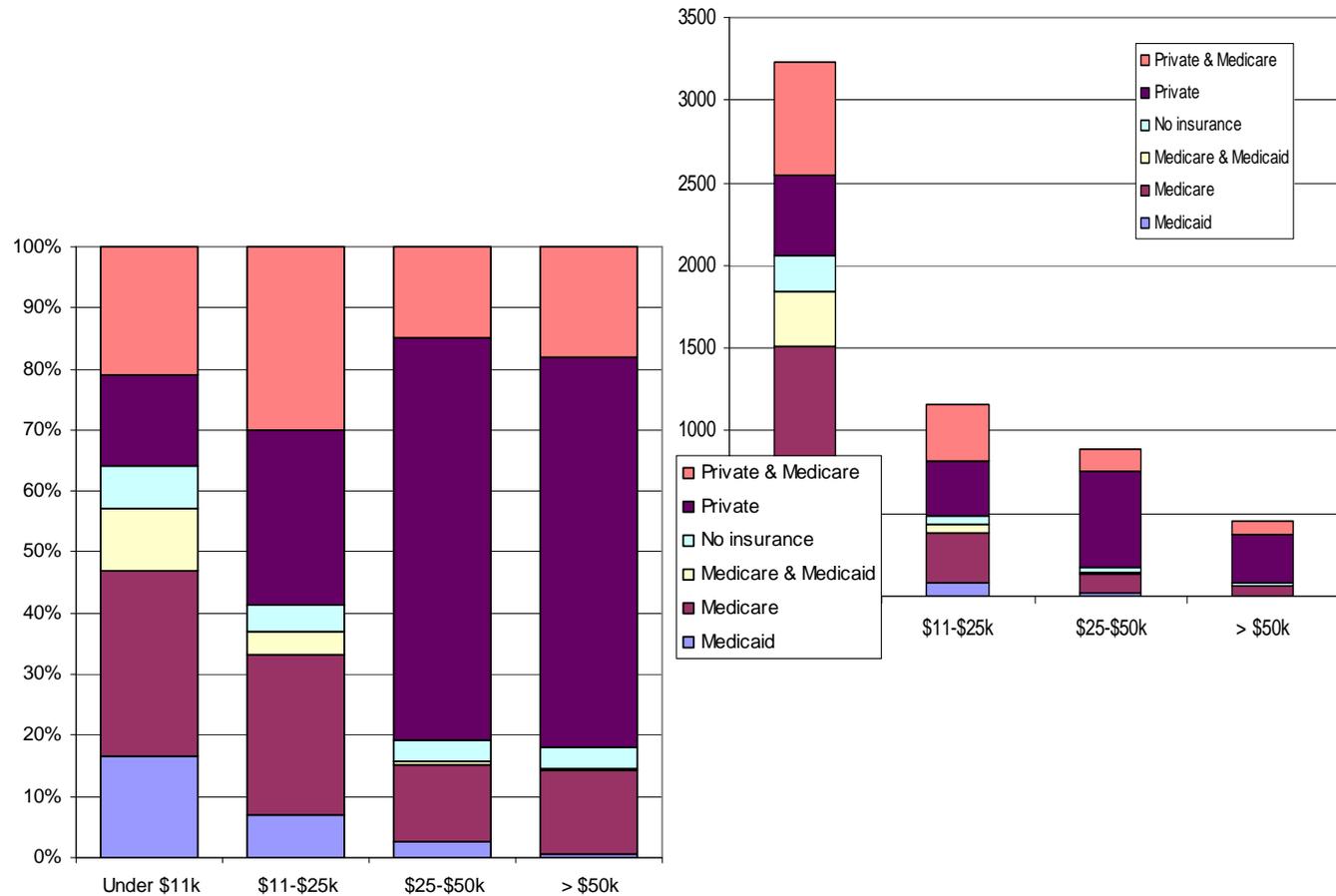
>



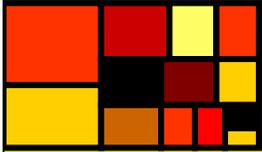
Cas bivarié distribution(qualitatif * qualitatif)



>



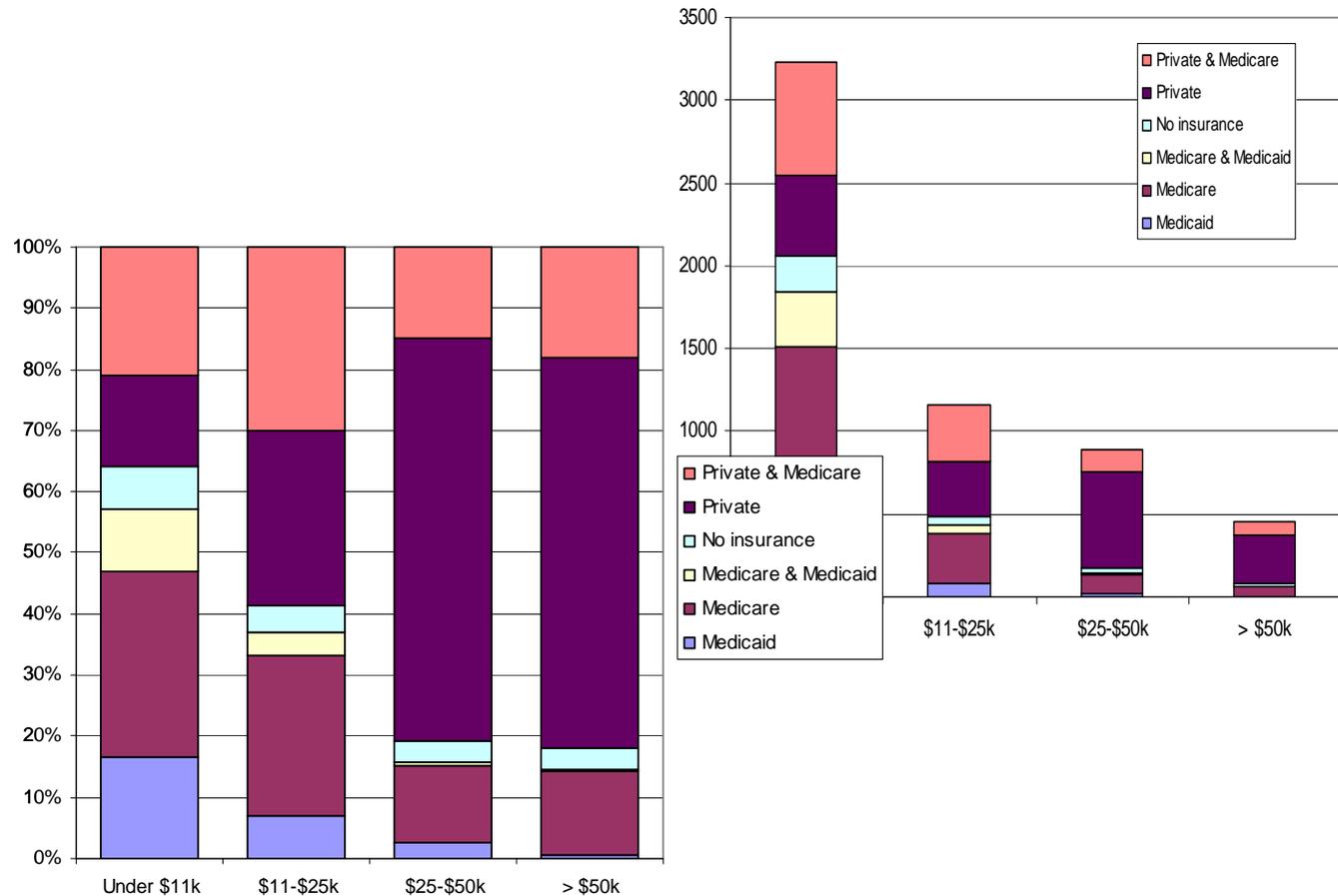
Cas bivarié distribution(qualitatif * qualitatif)



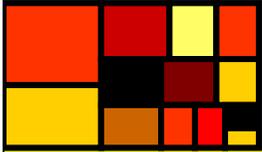
>

Les barres empilées sont une bonne solution sur Excel, chacune avec ses défauts :

- barre empilées en quantité : proportions intra-groupe peu visibles
- Barre empilées en proportion : perte de l'équilibre entre les groupes

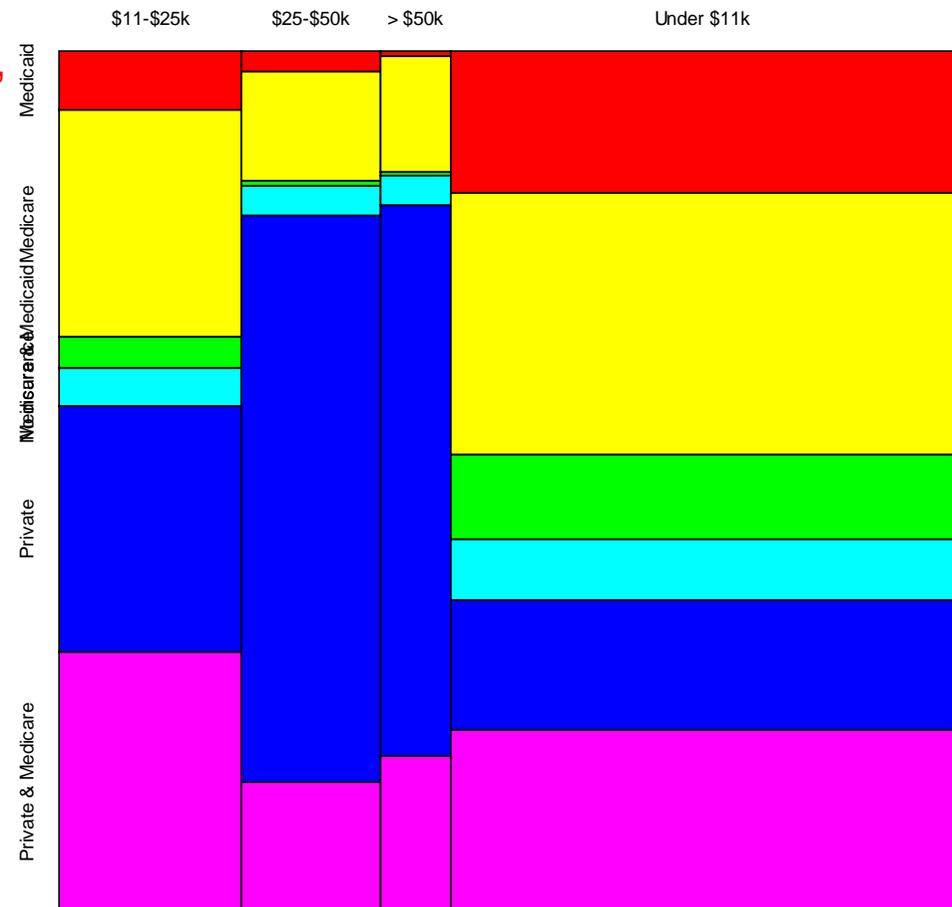


Cas bivarié distribution(qualitatif * qualitatif)

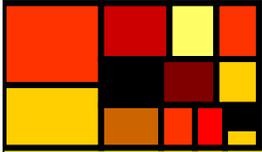


x

```
> plot(table(tableau$revenus, tableau$assurance),  
col=rainbow(6), off=0)
```

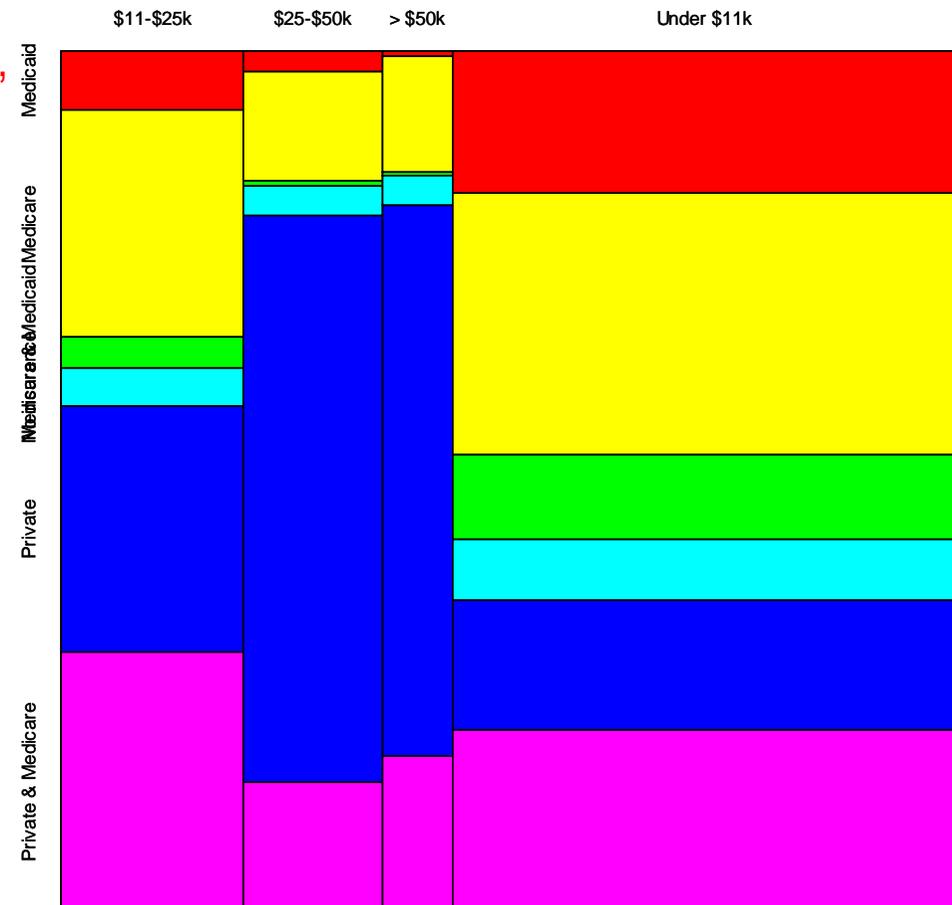


Cas bivarié distribution(qualitatif * qualitatif)

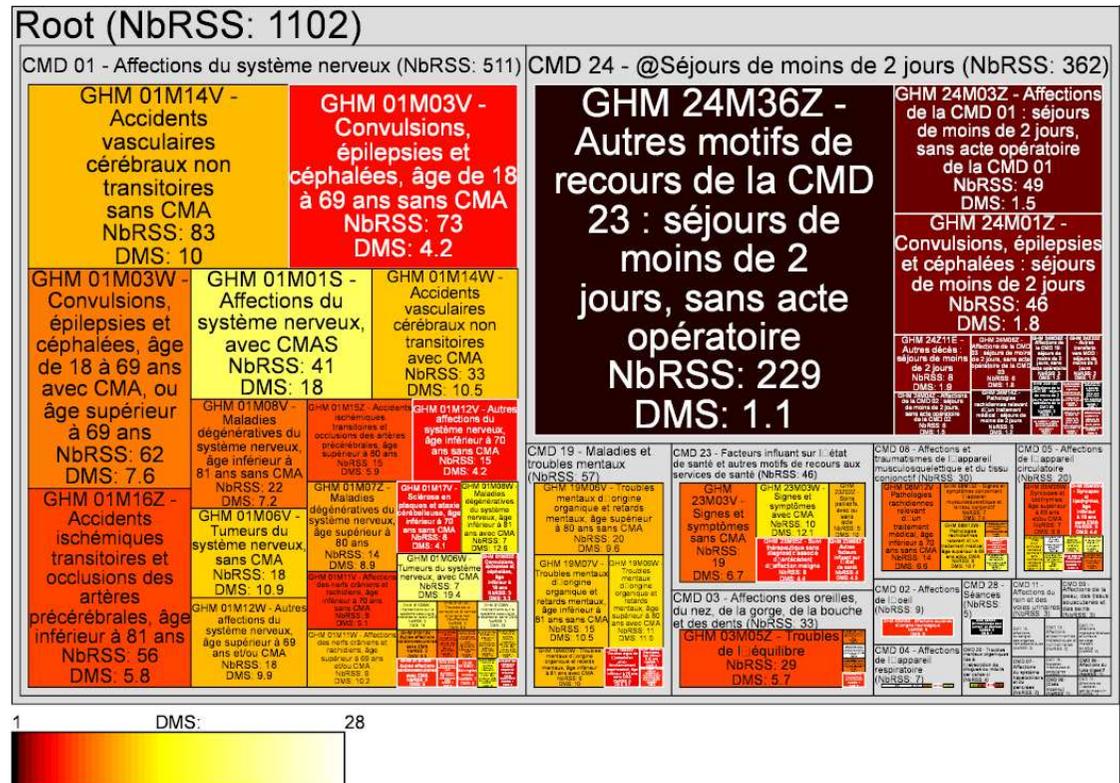
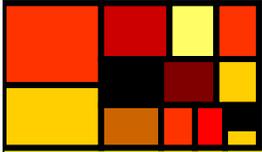


```
> plot(table(tableau$revenus, tableau$assurance),  
col=rainbow(6), off=0)
```

Le MosaicPlot de R est une solution parfaite aux problèmes de distribution jointe de deux variables qualitatives

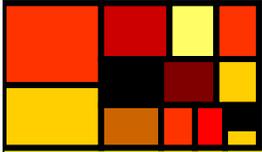


Cas bivarié distribution(qualitatif * qualitatif)



Cas bivarié

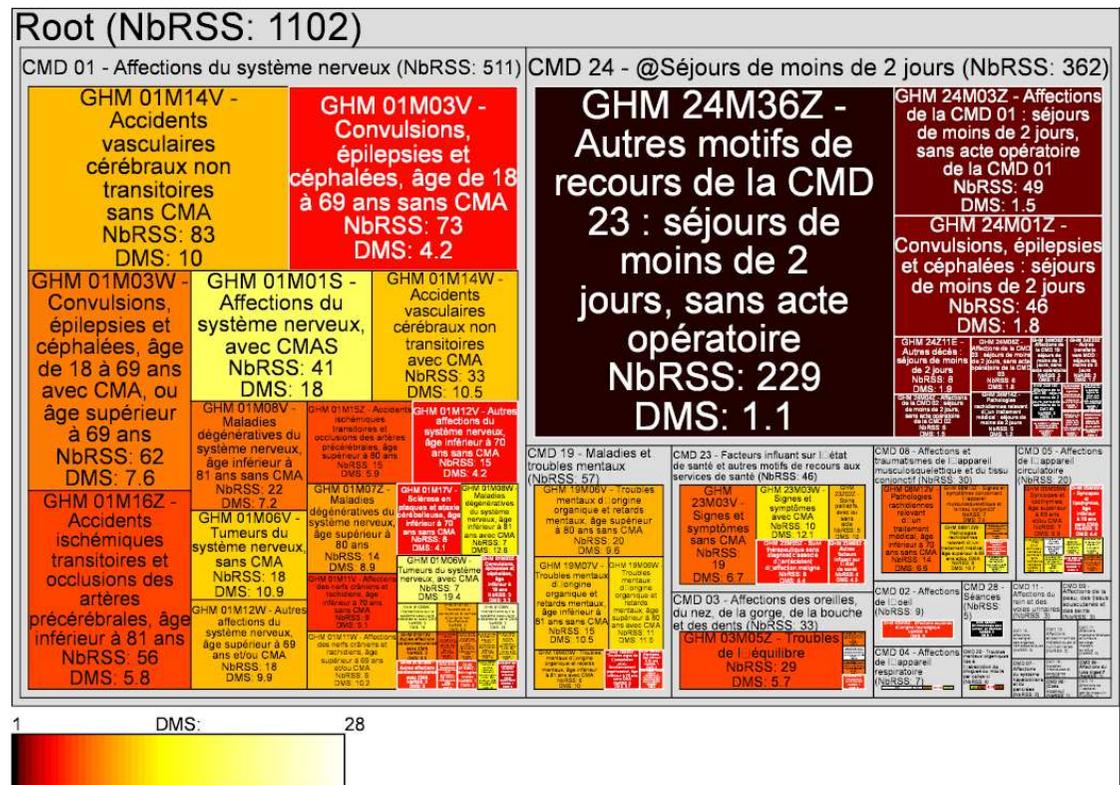
distribution(qualitatif * qualitatif)



Les TREEMAPS sont la solution idéale à ce type de problème.

Notes :

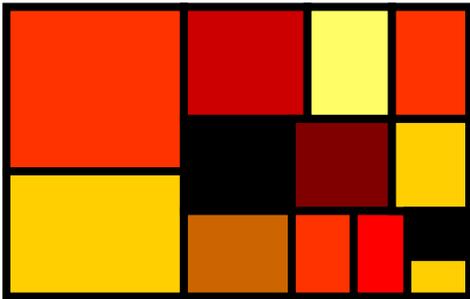
- il y a un choix délibéré d'imbrication des variables, contrairement au Mosaicplot.
- On peut donc l'utiliser pour redécouper des classes en sous-classes
- Affichage optimisé du texte
- Il est également possible d'utiliser la couleur pour représenter des données
- On peut utiliser la taille des rectangles pour autre chose que les effectifs



Cas pratiques – graphiques licites

Pour chaque cas, précisez :

- i. Le type de graphique
- ii. Les variables impliquées, leur type
- iii. Le cas :
Données natives
Distribution
Agrégation



Root (NbRSS: 1102)

CMD 01 - Affections du système nerveux (NbRSS: 511)

CMD 24 - @Séjours de moins de 2 jours (NbRSS: 362)

GHM 01M14V - Accidents vasculaires cérébraux non transitoires sans CMA NbRSS: 83 DMS: 10		GHM 01M03V - Convulsions, épilepsies et céphalées, âge de 18 à 69 ans sans CMA NbRSS: 73 DMS: 4.2		
GHM 01M03W - Convulsions, épilepsies et céphalées, âge de 18 à 69 ans avec CMA, ou âge supérieur à 69 ans NbRSS: 62 DMS: 7.6	GHM 01M01S - Affections du système nerveux, avec CMAS NbRSS: 41 DMS: 18		GHM 01M14W - Accidents vasculaires cérébraux non transitoires avec CMA NbRSS: 33 DMS: 10.5	
	GHM 01M08V - Maladies dégénératives du système nerveux, âge inférieur à 81 ans sans CMA NbRSS: 22 DMS: 7.2	GHM 01M15Z - Accidents transitoires et occlusions des artères précérébrales, âge supérieur à 80 ans NbRSS: 15 DMS: 5.9	GHM 01M12V - Autres affections du système nerveux, âge inférieur à 70 ans sans CMA NbRSS: 15 DMS: 4.2	
	GHM 01M16Z - Accidents ischémiques transitoires et occlusions des artères précérébrales, âge inférieur à 81 ans NbRSS: 56 DMS: 5.8	GHM 01M07Z - Maladies dégénératives du système nerveux, âge supérieur à 80 ans NbRSS: 14 DMS: 8.9	GHM 01M17V - Sclérose en plaques et ataxie cérébelleuse, âge inférieur à 70 ans sans CMA NbRSS: 8 DMS: 4.1	GHM 01M08W - Maladies dégénératives du système nerveux, âge inférieur à 81 ans avec CMA NbRSS: 7 DMS: 12.6
GHM 01M06V - Tumeurs du système nerveux, sans CMA NbRSS: 18 DMS: 10.9		GHM 01M11V - Affections des nerfs crâniens et rachidiens, âge inférieur à 70 ans sans CMA NbRSS: 9 DMS: 9.1	GHM 01M06W - Tumeurs du système nerveux, avec CMA NbRSS: 7 DMS: 19.4	GHM 01M09W - Troubles mentaux d'origine organique et retards mentaux, âge supérieur à 80 ans sans CMA NbRSS: 20 DMS: 9.6
GHM 01M12W - Autres affections du système nerveux, âge supérieur à 69 ans et/ou CMA NbRSS: 18 DMS: 9.9		GHM 01M11W - Affections des nerfs crâniens et rachidiens, âge supérieur à 69 ans et/ou CMA NbRSS: 9 DMS: 10.2	GHM 01M07W - Troubles mentaux d'origine organique et retards mentaux, âge inférieur à 81 ans sans CMA NbRSS: 15 DMS: 10.5	GHM 01M09W - Troubles mentaux d'origine organique et retards mentaux, âge supérieur à 80 ans avec CMA NbRSS: 11 DMS: 11.6

GHM 24M36Z -
 Autres motifs de recours de la CMD 23 : séjours de moins de 2 jours, sans acte opératoire
 NbRSS: 229
 DMS: 1.1

GHM 24M03Z - Affections de la CMD 01 : séjours de moins de 2 jours, sans acte opératoire de la CMD 01 NbRSS: 49 DMS: 1.5	GHM 24M01Z - Convulsions, épilepsies et céphalées : séjours de moins de 2 jours NbRSS: 46 DMS: 1.8
GHM 24Z11E - Autres décès : séjours de moins de 2 jours, sans acte opératoire de la CMD 01 NbRSS: 8 DMS: 1.9	GHM 24M05Z - Affections de la CMD 03 : séjours de moins de 2 jours, sans acte opératoire de la CMD 03 NbRSS: 8 DMS: 1.8

CMD 19 - Maladies et troubles mentaux (NbRSS: 57)

GHM 19M08V - Troubles mentaux d'origine organique et retards mentaux, âge supérieur à 80 ans sans CMA NbRSS: 20 DMS: 9.6	GHM 19M07V - Troubles mentaux d'origine organique et retards mentaux, âge inférieur à 81 ans sans CMA NbRSS: 15 DMS: 10.5
GHM 19M09V - Troubles mentaux d'origine organique et retards mentaux, âge supérieur à 80 ans avec CMA NbRSS: 11 DMS: 11.6	GHM 19M08W - Troubles mentaux d'origine organique et retards mentaux, âge supérieur à 81 ans avec CMA NbRSS: 15 DMS: 10.5

CMD 23 - Facteurs influant sur l'état de santé et autres motifs de recours aux services de santé (NbRSS: 46)

GHM 23M03V - Signes et symptômes sans CMA NbRSS: 19 DMS: 6.7	GHM 23M03W - Signes et symptômes avec CMA NbRSS: 10 DMS: 12.1
--	---

CMD 03 - Affections des oreilles, du nez, de la gorge, de la bouche et des dents (NbRSS: 33)

GHM 03M05Z - Troubles de l'équilibre NbRSS: 29 DMS: 5.7

CMD 08 - Affections et traumatismes de l'appareil musculo-squelettique et du tissu conjonctif (NbRSS: 30)

GHM 08M12V - Pathologies rachidiennes relevant d'un traitement médical, âge inférieur à 70 ans sans CMA NbRSS: 14 DMS: 9.6
--

CMD 02 - Affections de l'oeil (NbRSS: 9)

CMD 28 - Séances (NbRSS: 5)

CMD 11 - Affections du rein et des voies urinaires (NbRSS: 3)

CMD 09 - Affections de la peau, des yeux, des oreilles et des dents (NbRSS: 3)

CMD 04 - Affections de l'appareil respiratoire (NbRSS: 7)

CMD 05 - Affections de l'appareil circulatoire (NbRSS: 20)

CMD 06 - Affections de l'appareil digestif (NbRSS: 10)

CMD 07 - Affections de l'appareil génital (NbRSS: 10)

CMD 10 - Affections de l'appareil locomoteur (NbRSS: 10)

CMD 12 - Affections de l'appareil urinaire (NbRSS: 10)

CMD 13 - Affections de l'appareil respiratoire (NbRSS: 10)

CMD 14 - Affections de l'appareil digestif (NbRSS: 10)

CMD 15 - Affections de l'appareil génital (NbRSS: 10)

CMD 16 - Affections de l'appareil locomoteur (NbRSS: 10)

CMD 17 - Affections de l'appareil urinaire (NbRSS: 10)

CMD 18 - Affections de l'appareil respiratoire (NbRSS: 10)

CMD 19 - Affections de l'appareil digestif (NbRSS: 10)

CMD 20 - Affections de l'appareil génital (NbRSS: 10)

CMD 21 - Affections de l'appareil locomoteur (NbRSS: 10)

CMD 22 - Affections de l'appareil urinaire (NbRSS: 10)

CMD 23 - Affections de l'appareil respiratoire (NbRSS: 10)

CMD 24 - Affections de l'appareil digestif (NbRSS: 10)

CMD 25 - Affections de l'appareil génital (NbRSS: 10)

CMD 26 - Affections de l'appareil locomoteur (NbRSS: 10)

CMD 27 - Affections de l'appareil urinaire (NbRSS: 10)

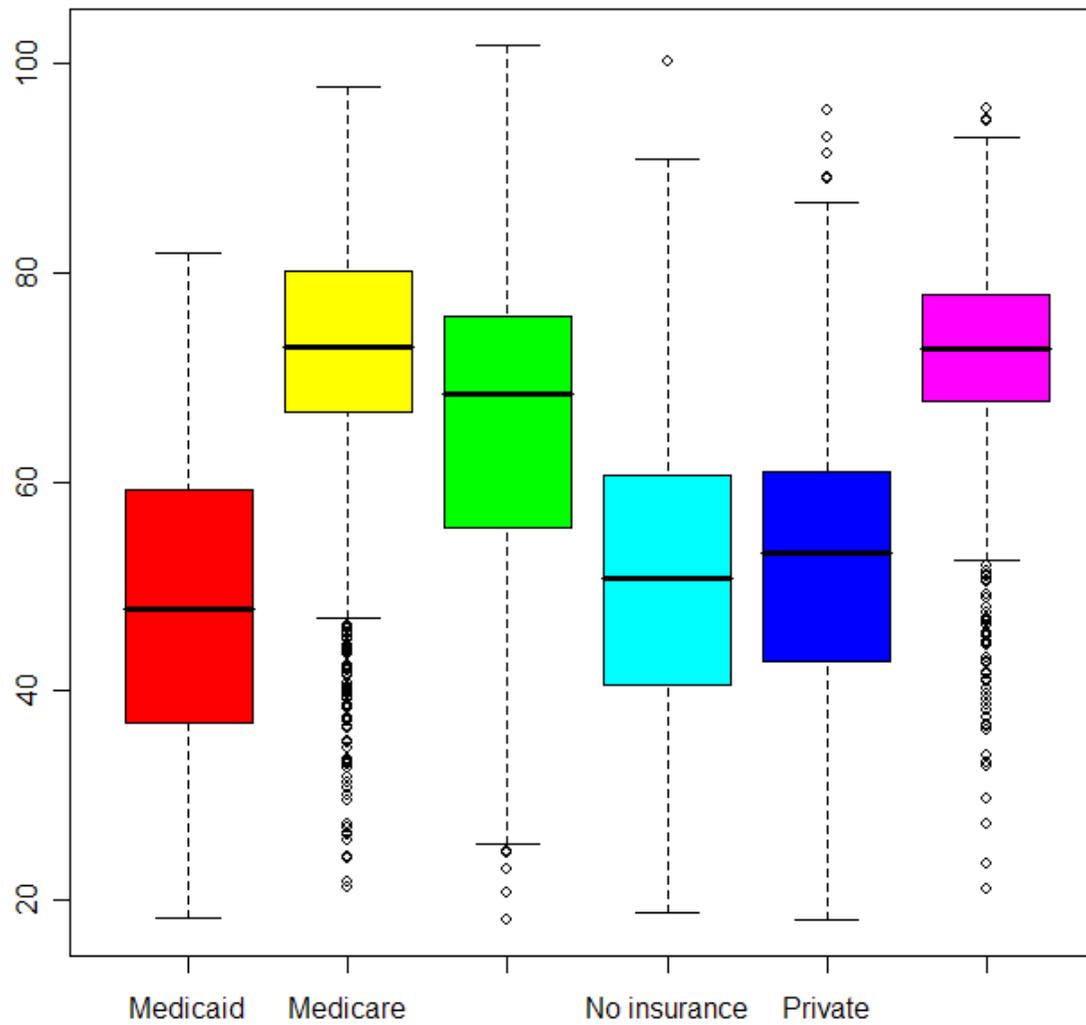
CMD 28 - Affections de l'appareil respiratoire (NbRSS: 10)

CMD 29 - Affections de l'appareil digestif (NbRSS: 10)

CMD 30 - Affections de l'appareil génital (NbRSS: 10)



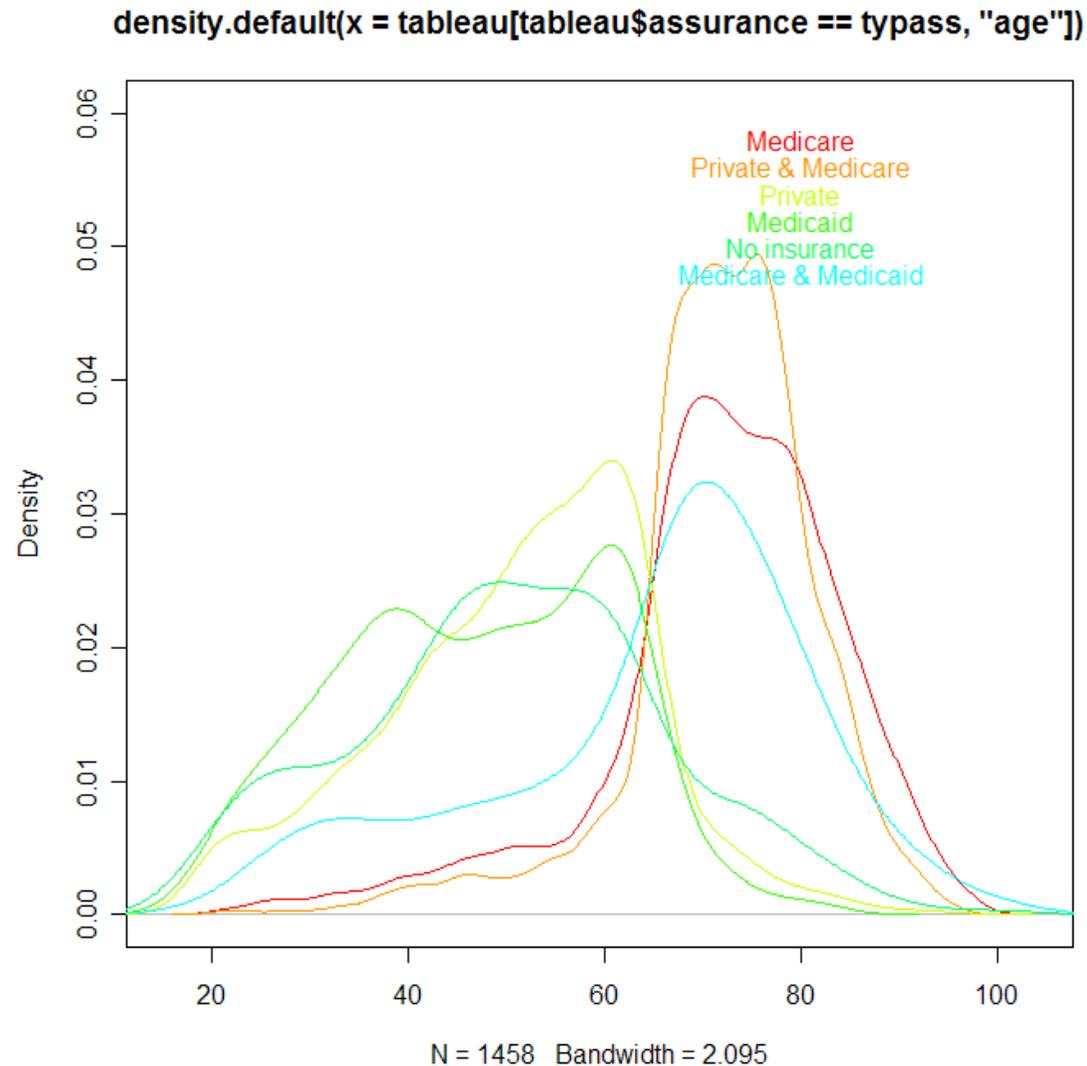
```
> boxplot( tableau$age ~ tableau$assurance, col=rainbow(6))
```



```

indice <- 1 ;
for( typass in as.vector(unique(tableau$assurance)) ) {
  couleur <- rainbow(10)[indice] ;
  if( indice==1 ) { plot(density(tableau[ tableau$assurance==typass ,"age"]), ylim=c(0,0.06), col=couleur) ; }
  else { lines(density(tableau[ tableau$assurance==typass ,"age"]), col=couleur) ; }
  text(x=80, y=0.06-0.002*indice, labels=typass, col=couleur) ;
  indice <- indice+1 ;
}

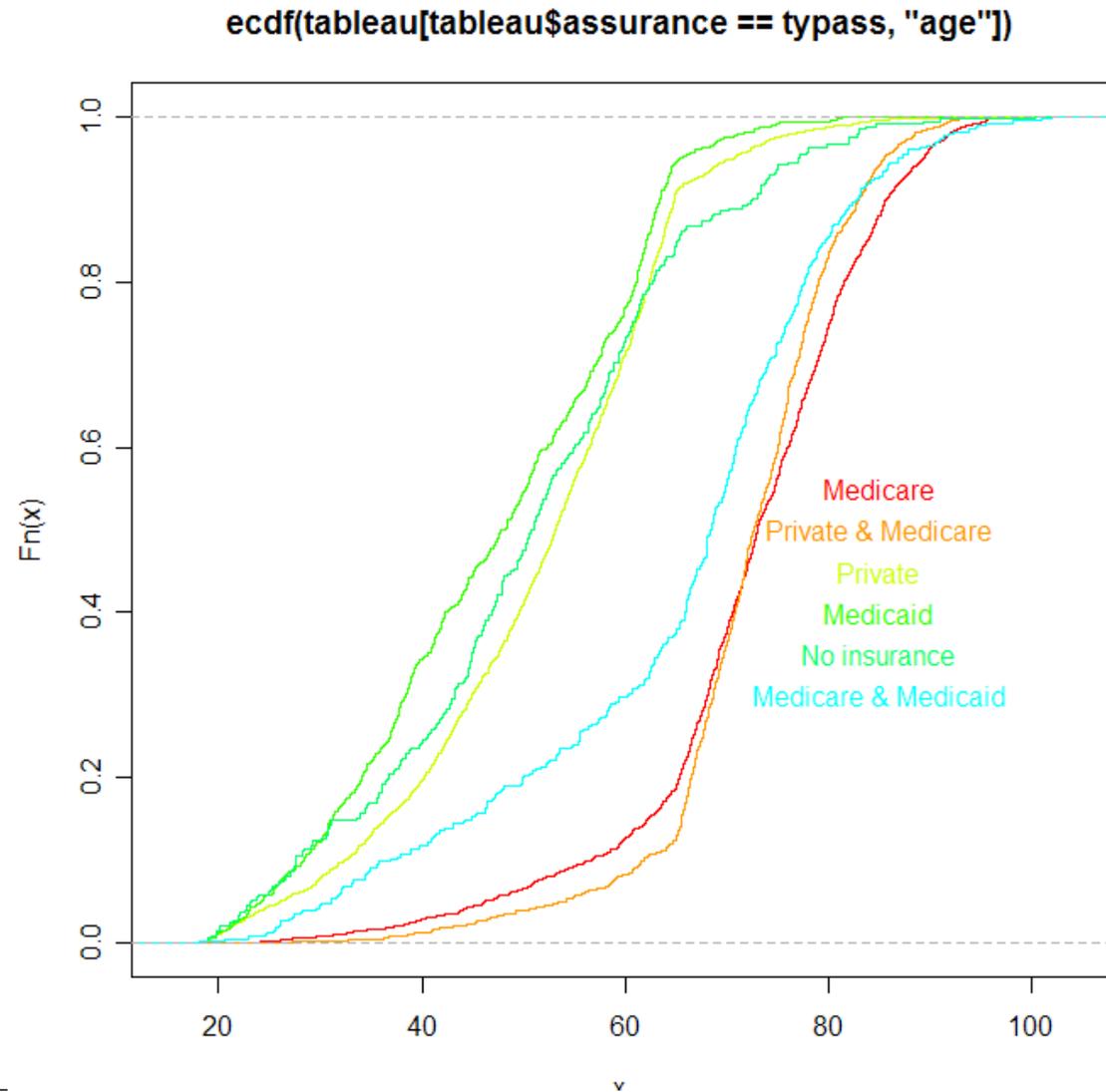
```



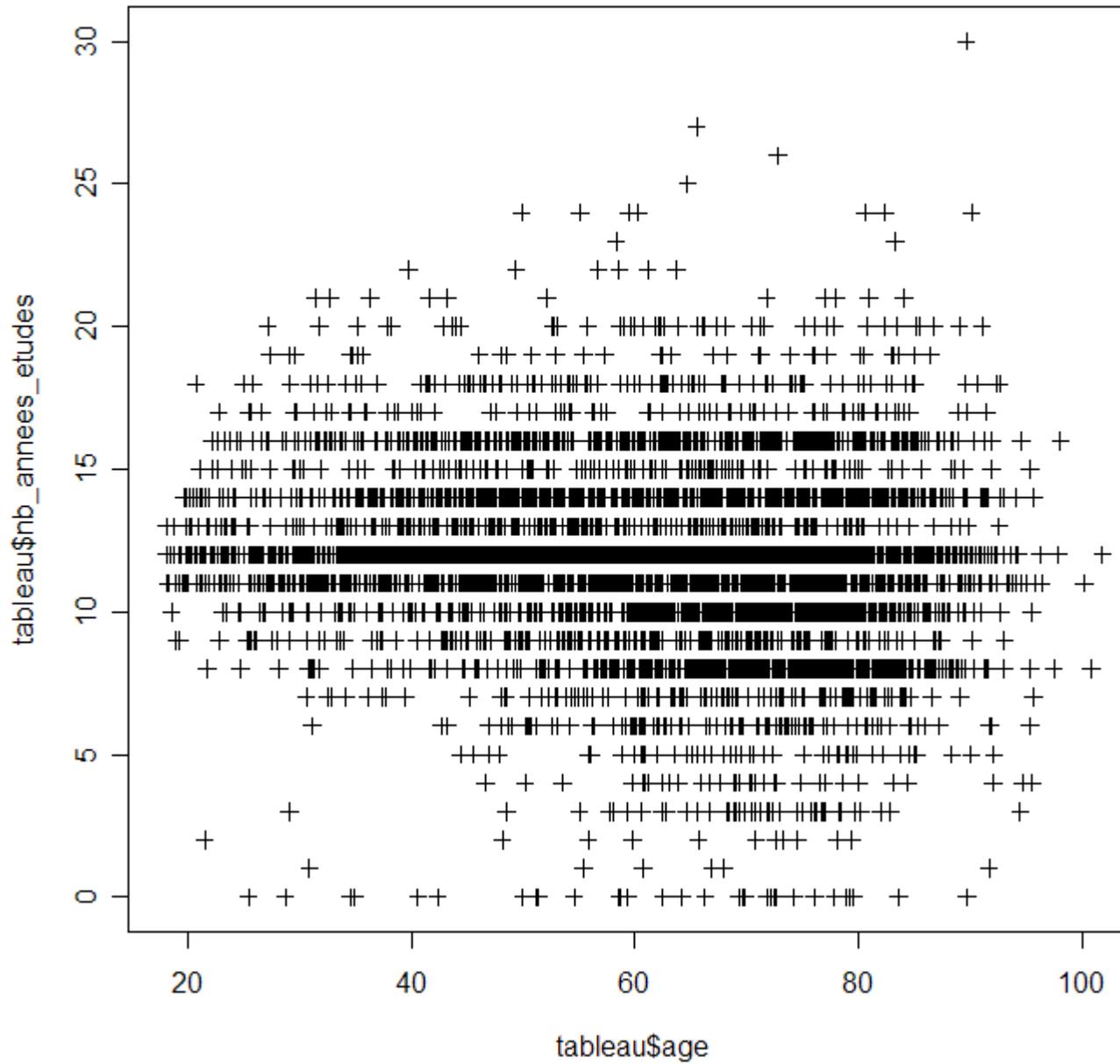
```

indice <- 1 ;
for( typass in as.vector(unique(tableau$assurance)) ) {
  couleur <- rainbow(10)[indice] ;
  if( indice==1 ) {
    plot(ecdf(tableau[ tableau$assurance==typass ,"age"]), verticals=TRUE, do.points=FALSE, col.hor=couleur, col.vert=couleur) ;
  } else {
    lines(ecdf(tableau[ tableau$assurance==typass ,"age"]), verticals=TRUE, do.points=FALSE, col.hor=couleur, col.vert=couleur) ;
  }
  text(x=85, y=0.6-0.05*indice, labels=typass, col=couleur) ;
  indice <- indice+1 ;
}

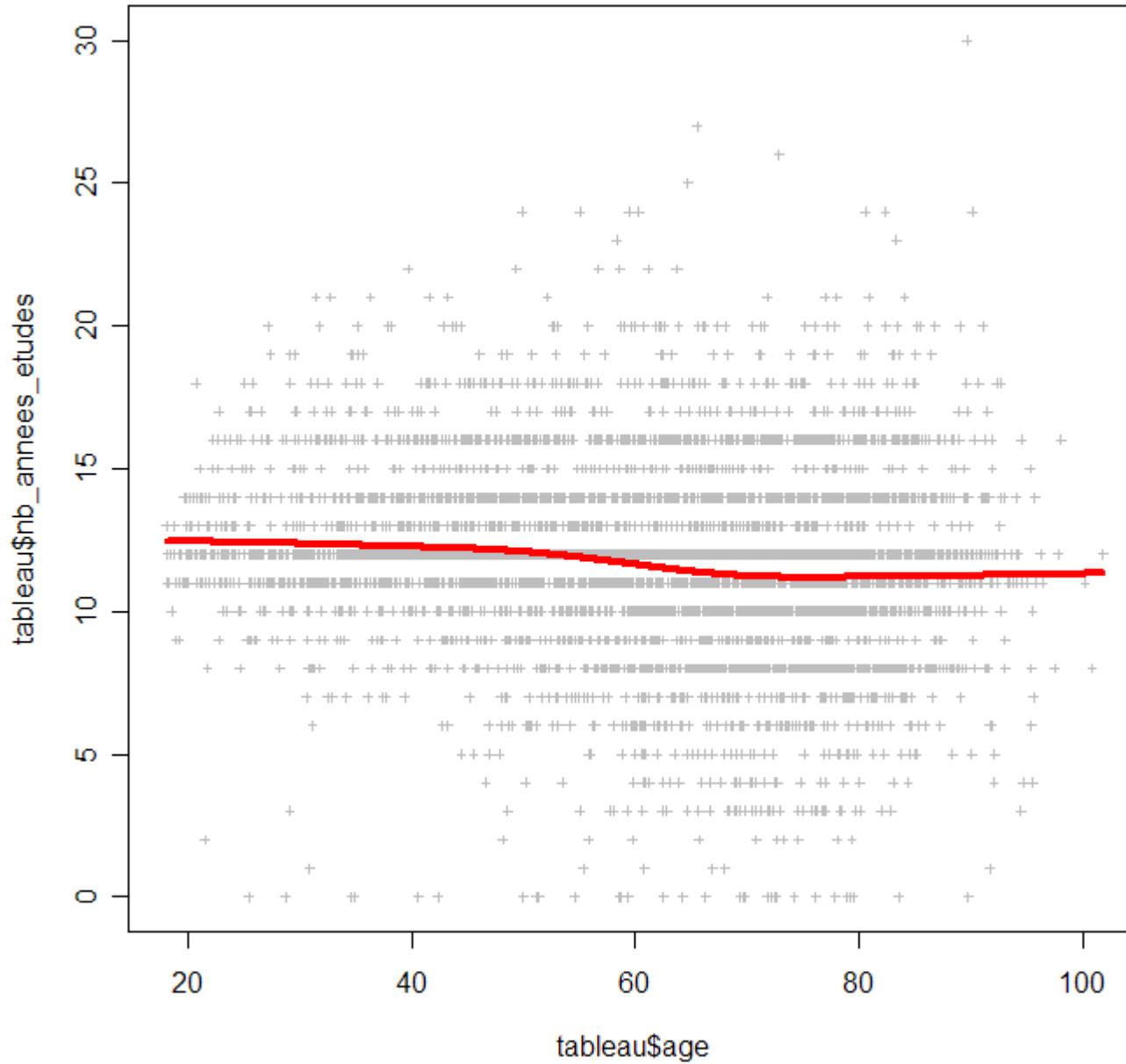
```



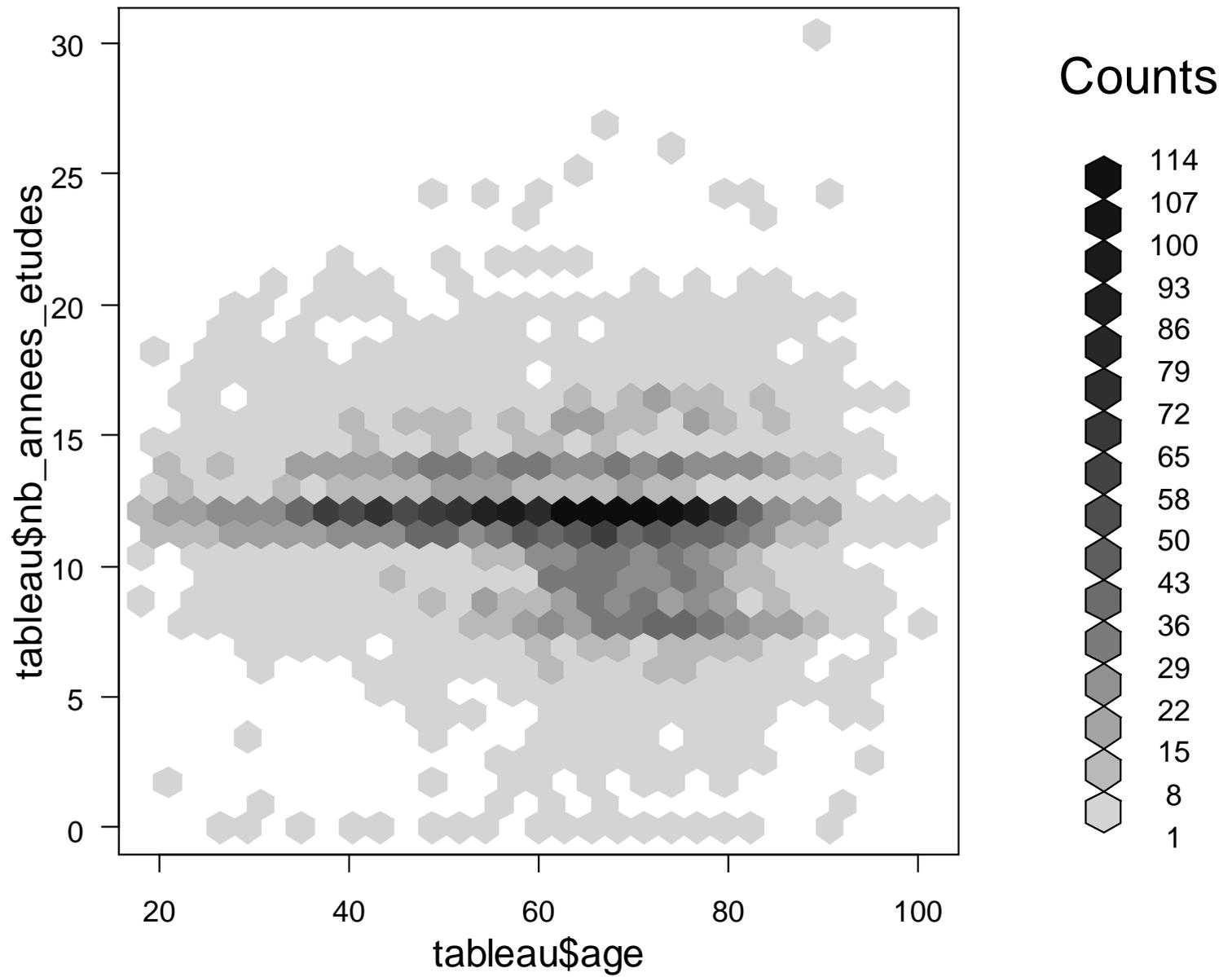
```
plot(x=tableau$age, y=tableau$nb_annees_etudes, pch=3) ;
```



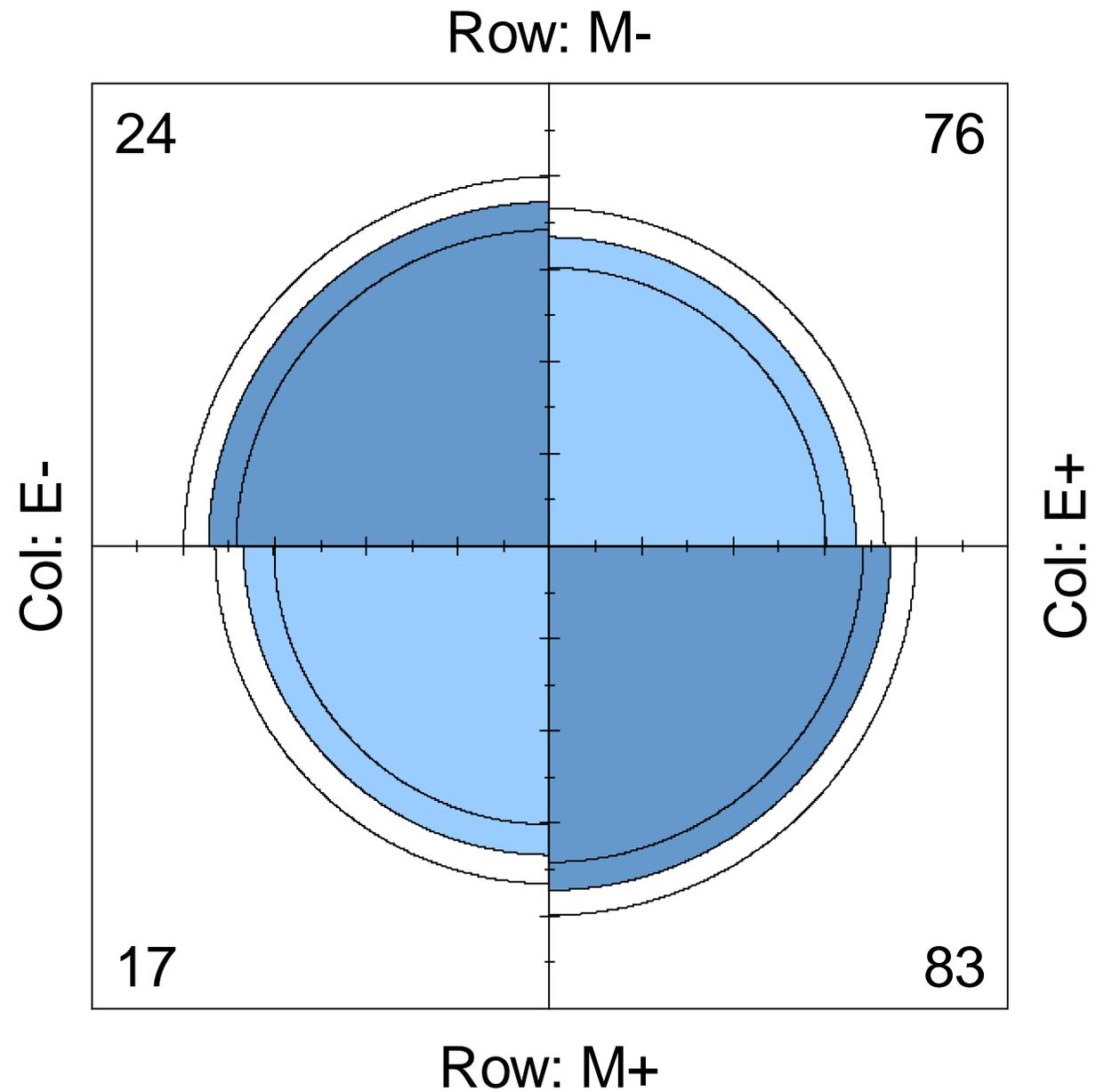
```
plot(x=tableau$age, y=tableau$nb_annees_etudes, pch=3, col="grey", cex=0.5) ;  
lines(lowess(x=tableau$age, y=tableau$nb_annees_etudes), col="red", lw=4) ;
```



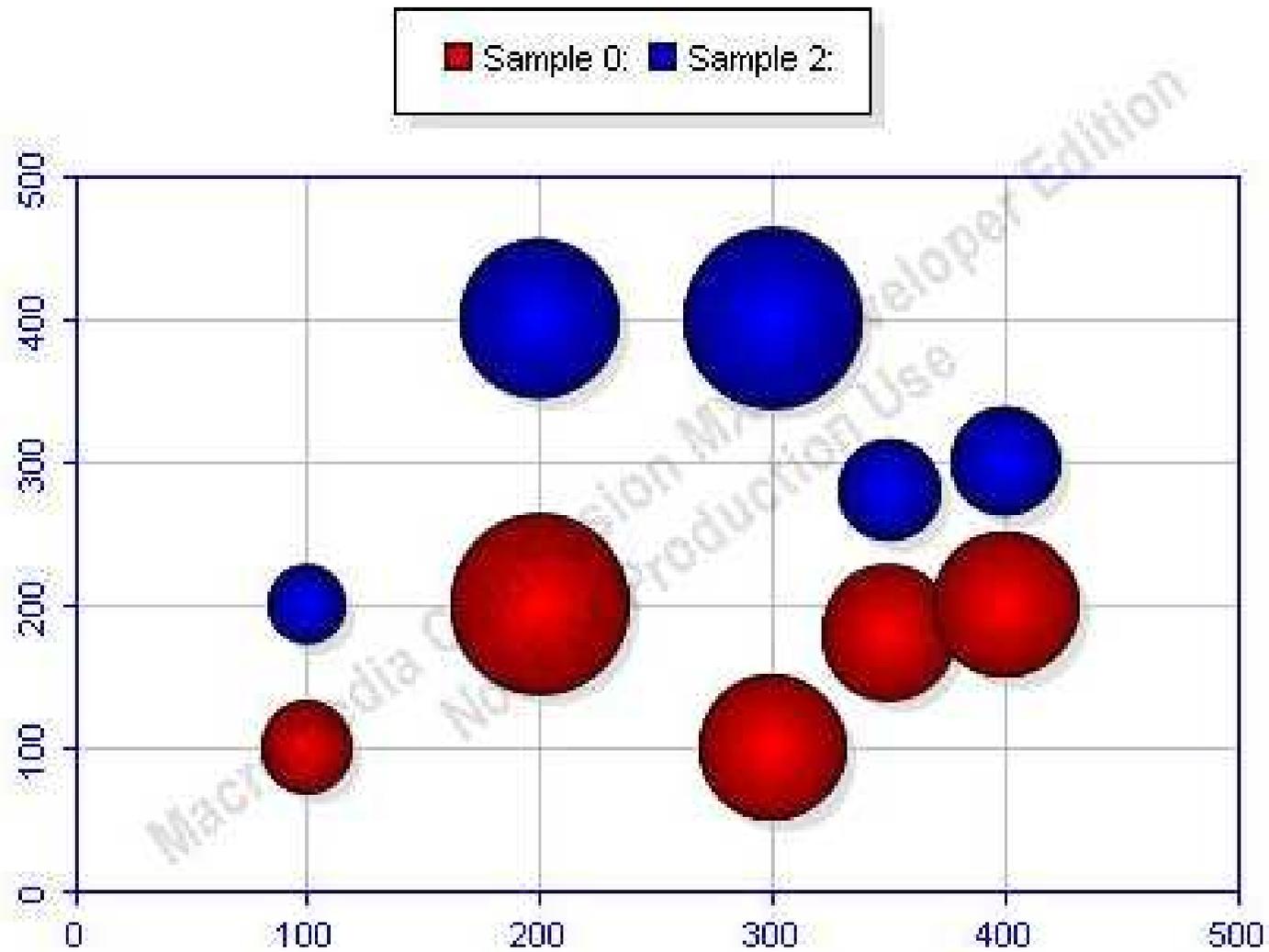
```
library(hexbin) ;  
plot(hexbin(x=tableau$age, y=tableau$nb_annees_etudes)) ;
```



```
x <- c(24,76,17,83)
a <- matrix(x,nr=2,byrow=T)
var1 <- c("E-","E+")
var2 <- c("M-","M+")
dimnames(a) <- list(var2,var1)
fourfoldplot(a)
```

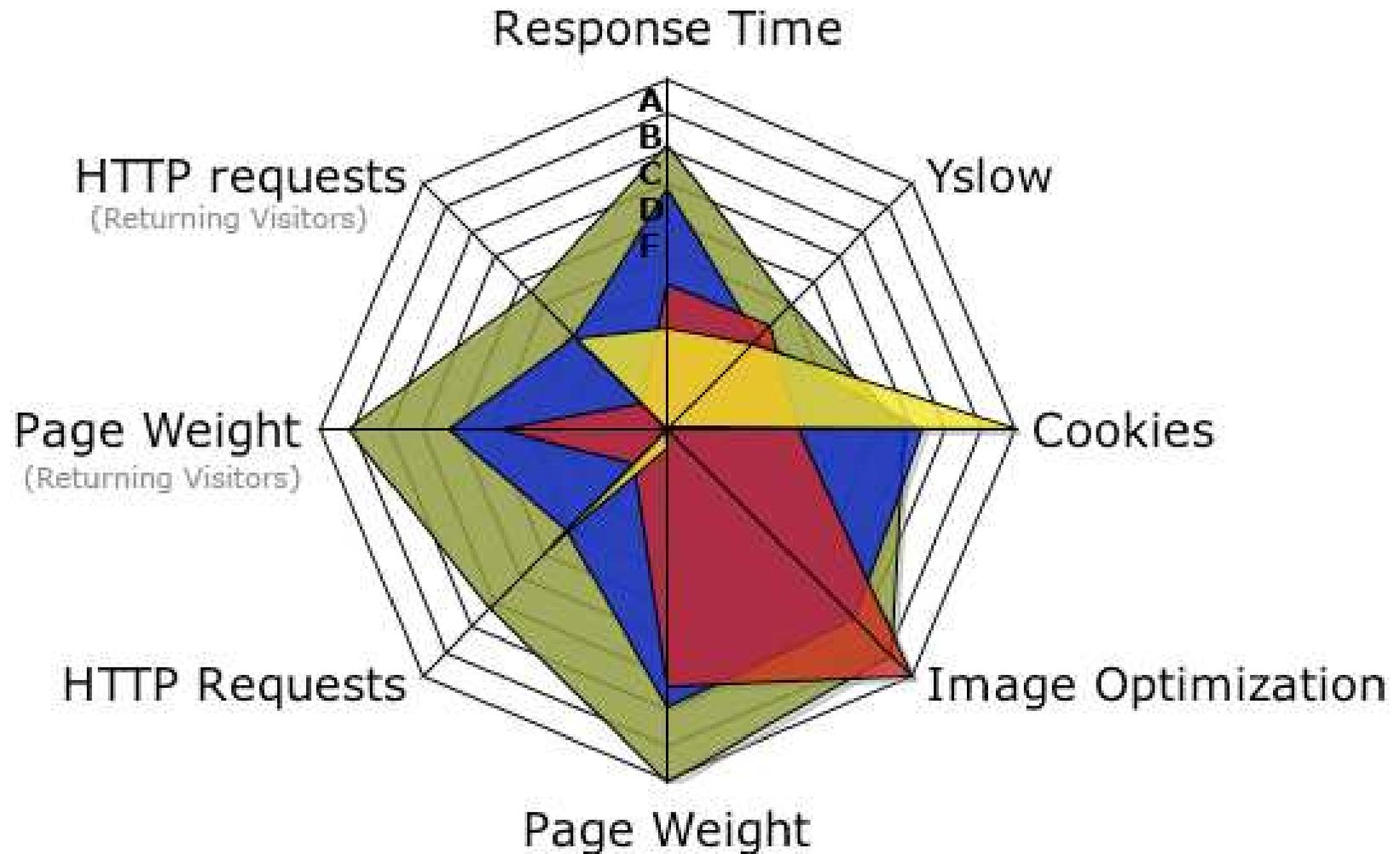


Bubble Chart

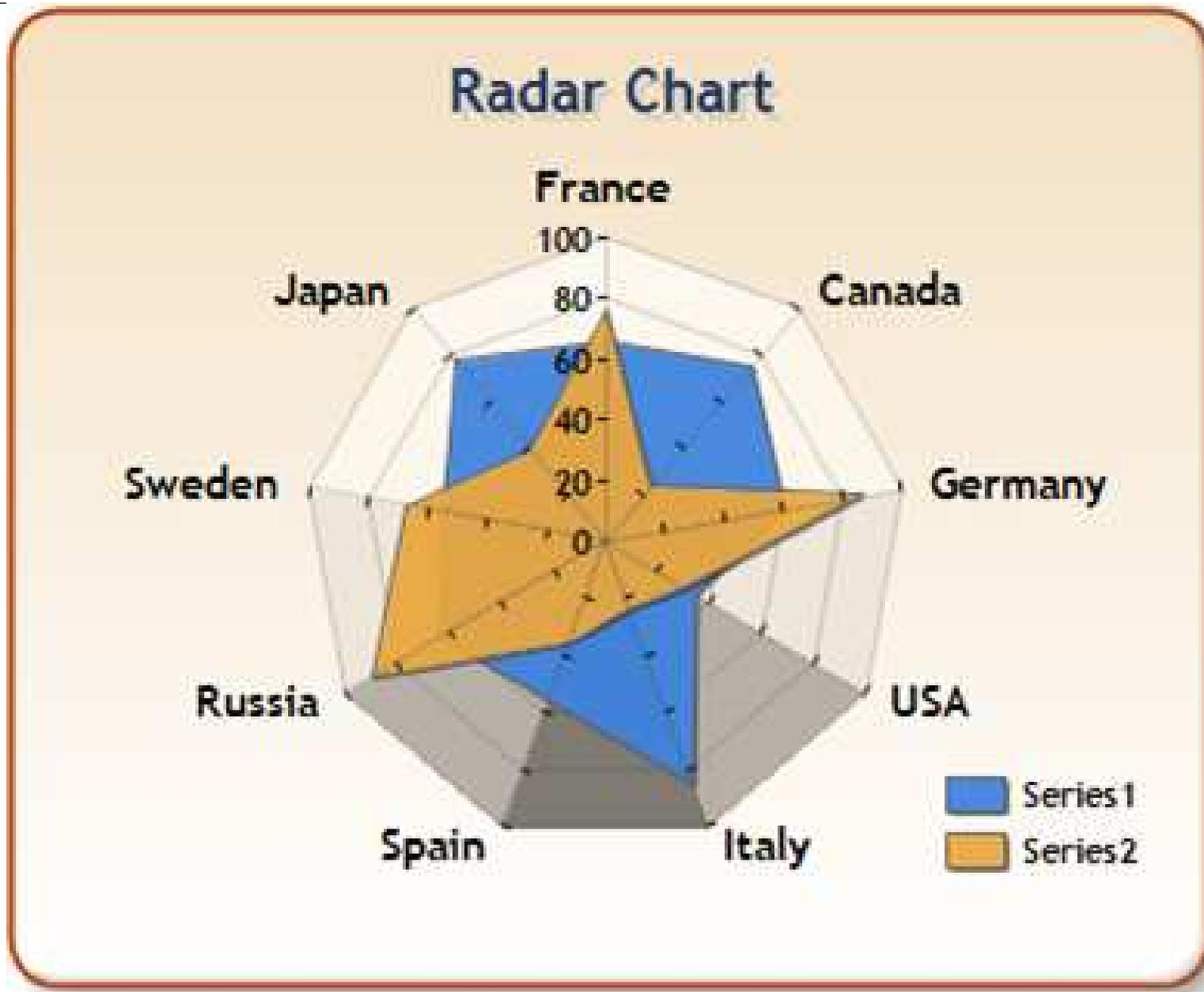


Republicans Report Card

■ Mike Huckabee ■ Ron Paul ■ John McCain ■ Mitt Romney



...



Exemple compatible : volumes de ventes (unité arbitraire) de deux produits dans différents pays