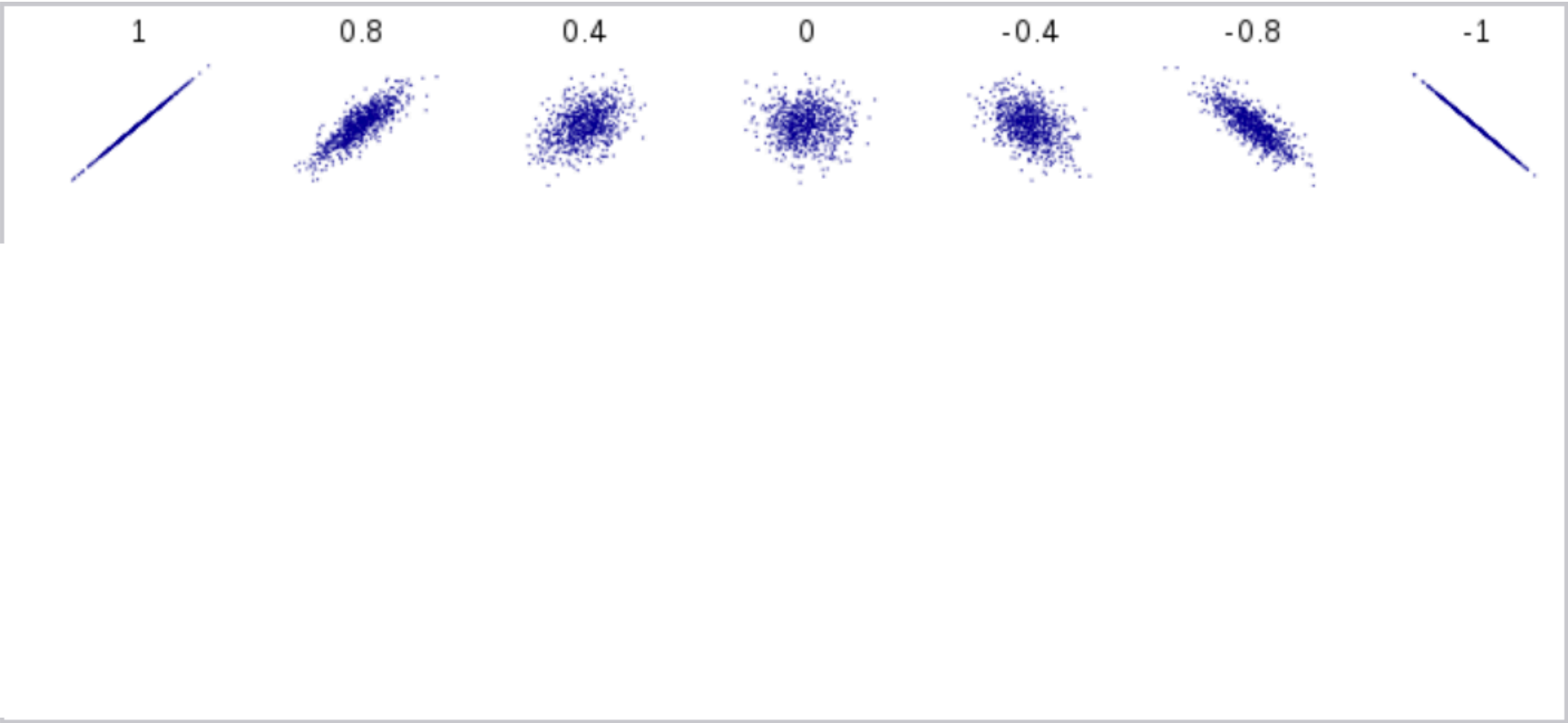
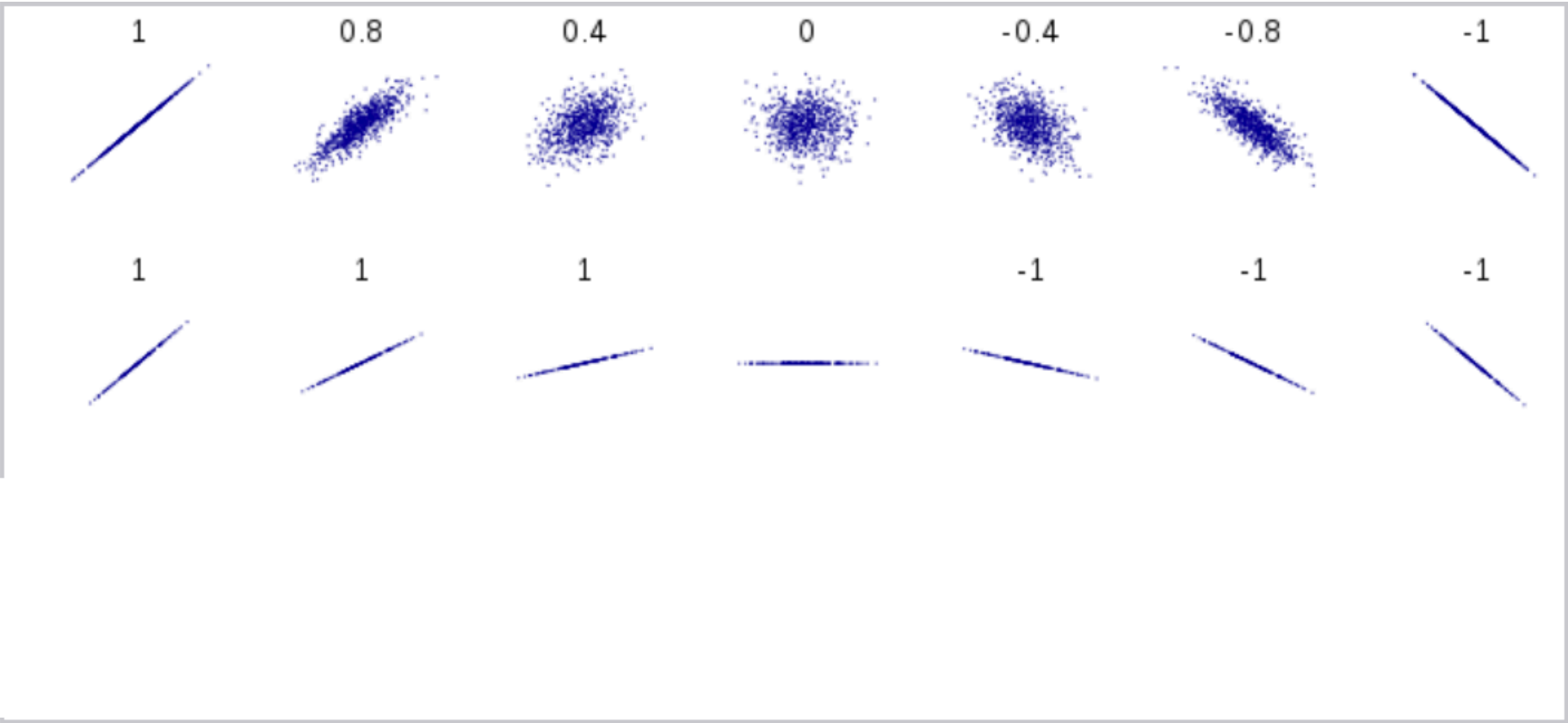
Café « Méthodes et Analyses des données »

Corrélations, seuils de validité, etc .. .

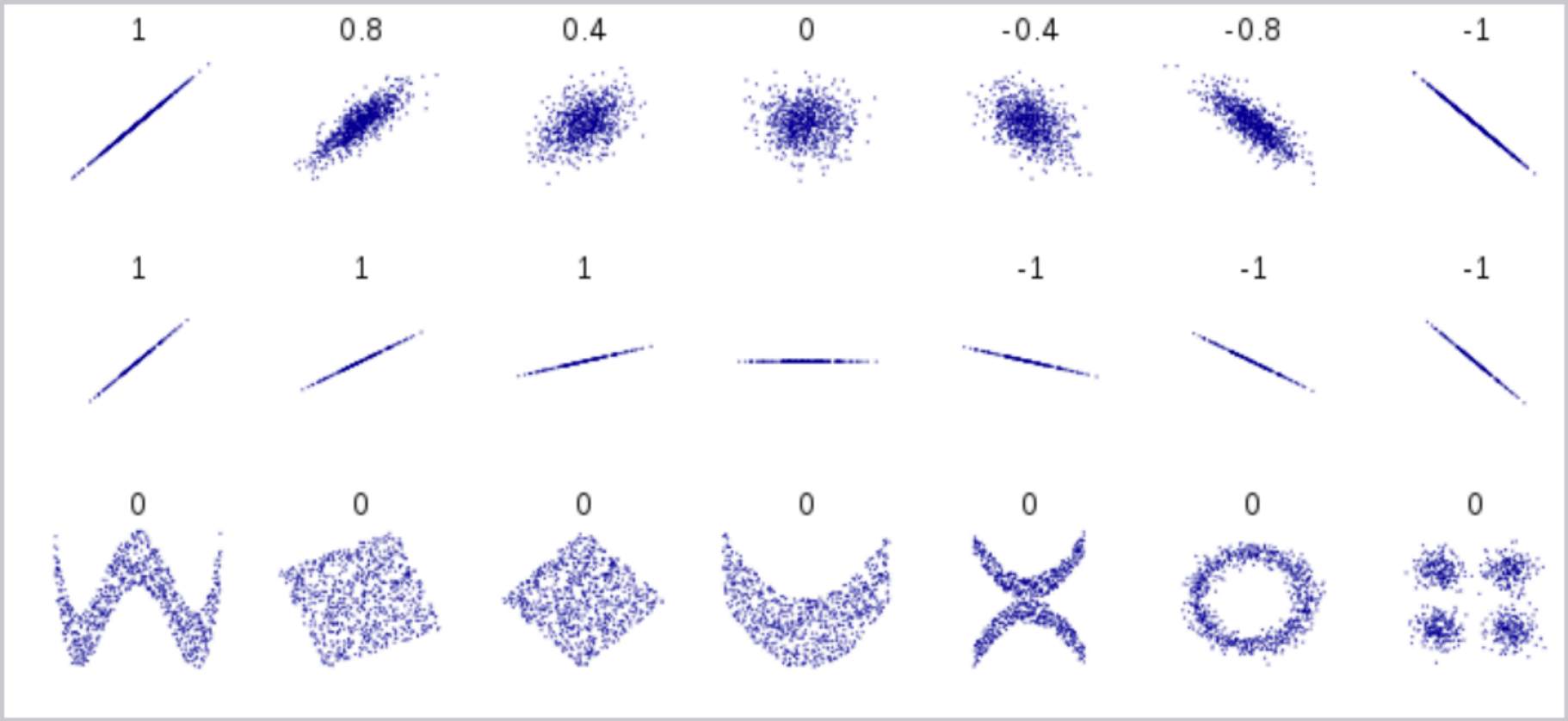
16 octobre 2019



Exemples de coefficients de corrélation : corrélation linéaire pour les deux premières lignes, non linéaire pour la troisième ligne.



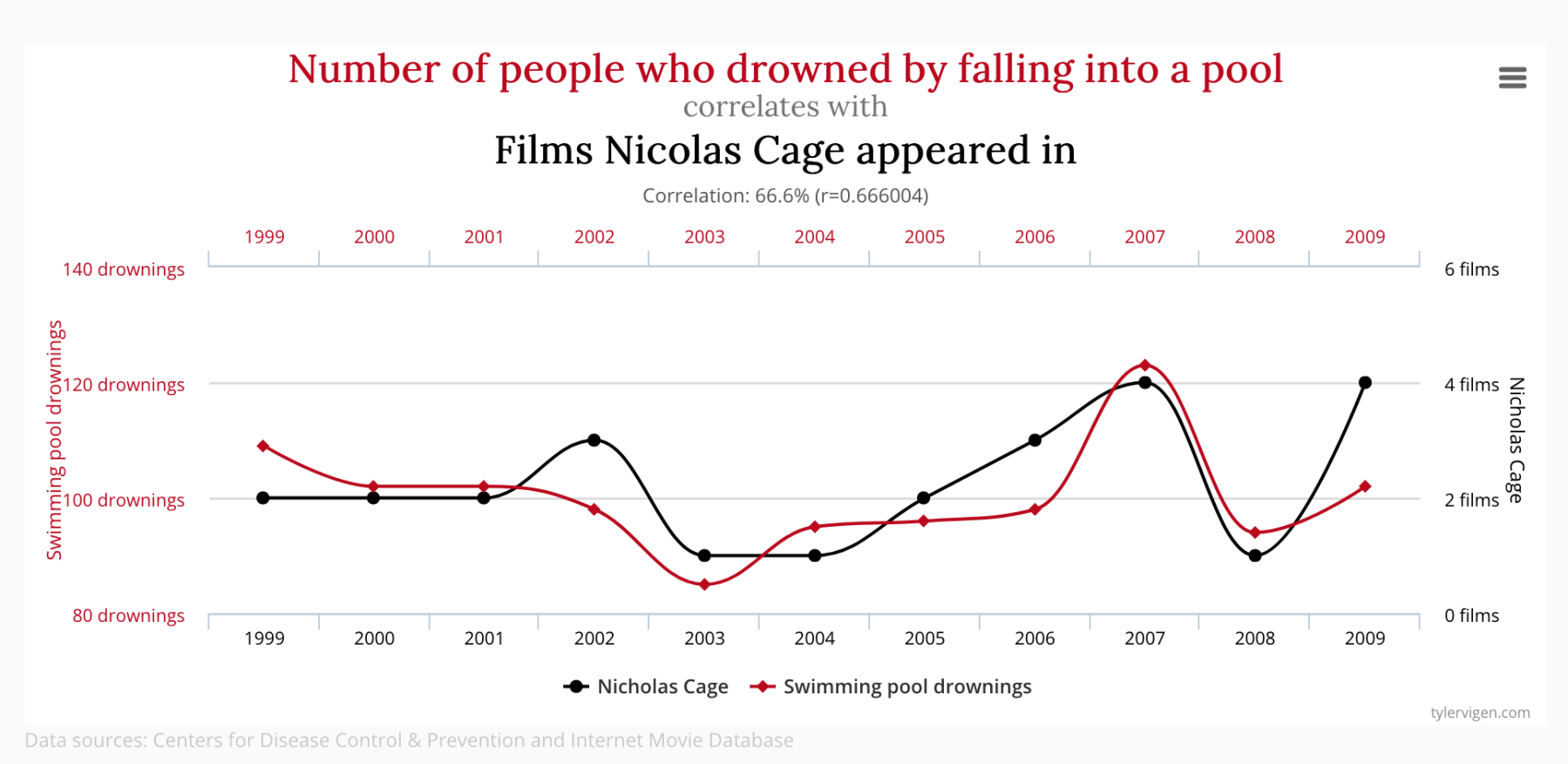
Exemples de coefficients de corrélation : corrélation linéaire pour les deux premières lignes, non linéaire pour la troisième ligne.



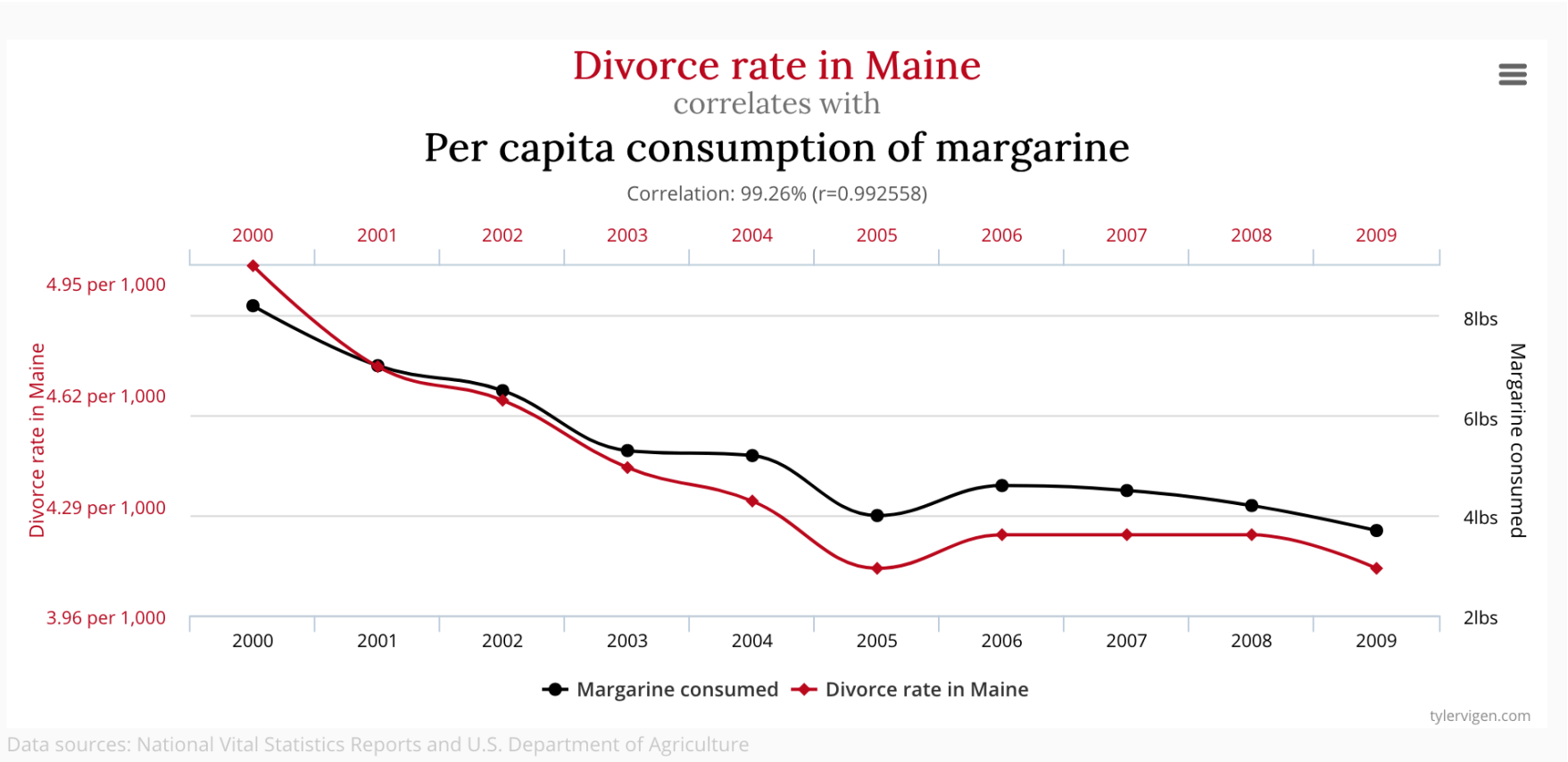
Exemples de coefficients de corrélation : corrélation linéaire pour les deux premières lignes, non linéaire pour la troisième ligne.

Corrélation et causalité

- Deux variables/mesures sont **<u>corrélées</u>** si
  - Elles évoluent dans le même sens (corrélation positive : quand l'un augmente, l'autre aussi)
  - Ou dans un sens opposé (corrélation négative : quand l'un augmente, l'autre diminue).
- Mais pas forcément de causalité entre ces deux variables/mesures



Source: http://tylervigen.com/spurious-correlations



Source: http://tylervigen.com/spurious-correlations

Corrélation et causalité

- Deux variables/mesures sont **corrélées** si
  - Elles évoluent dans le même sens (corrélation positive : quand l'un augmente, l'autre aussi)
  - Ou dans un sens opposé (corrélation négative : quand l'un augmente, l'autre diminue).
- Mais pas forcément de causalité entre ces deux variables/mesures

- La **corrélation** est donc un lien entre deux variables/ mesures , sans que pour autant la variation de l'une soit **la cause** de la variation de l'autre.

Variable latente

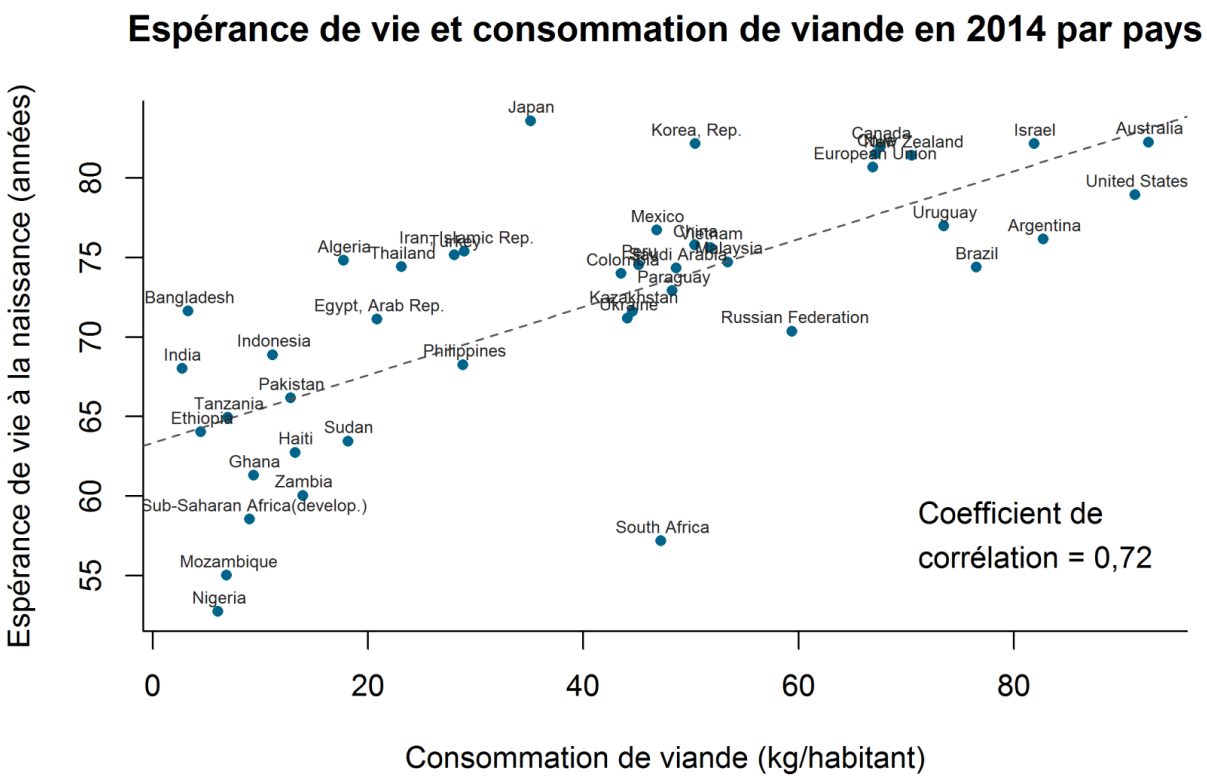


Figure 3 : Espérance de vie à la naissance et consommation de viande en 2014 dans certains pays du monde. Sources : OECD-FAO Agricultural Outlook (Edition 2015) et The World Bank, World Development Indicators.

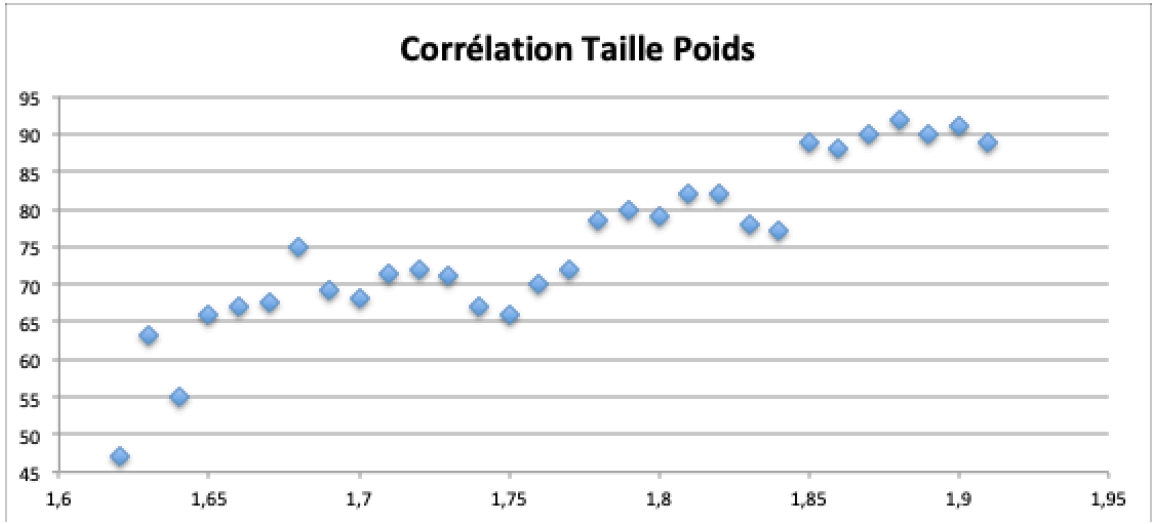
https://soepidemio.com/2016/11/14/correlation-statistique-prudence-a-linterpretation/

Corrélation et dépendances

- Taille et poids ?
  - Quanti*Quanti => R2
- Couleur des yeux et couleurs des cheveux ?
  - Quali*Quali => Chi2
- Couleur des cheveux et taille ?
  - Quali*Quanti =>
    - recoder le quanti en quali (classe de taille)
    - Faire des tests, comparaison de moyennes

Corrélation et Dépendance

- Corrélation : Variables quantitatives/ordinales
  - Graphes XY
  - R2, R de Pearson => variables normales
  - Coefficient de Spearman => non normalité, indicateurs non paramétriques
  - Exemple : Taille et poids
- Dépendance : Variables Qualitatives/nominales
  - Tableau de contingence
  - Calcul du Chi2
  - Exemple : Couleur des yeux et couleur des cheveux

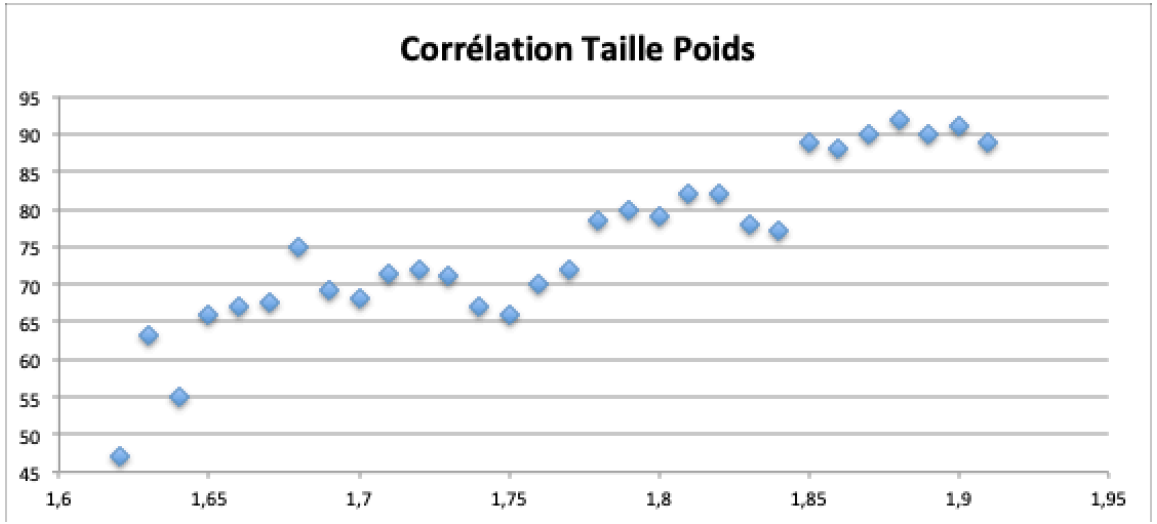


H0 : R=0
H1 : R≠0
R=0,91 - p value <0,05

	blonds	bruns	noirs	roux	Total
bleus	25	9	3	7	44
gris	13	17	10	7	47
marrons	7	13	8	5	33
Total	45	39	21	19	124

Répartition couleur cheveux/yeux

H0 : Variable A est indépendante de la variable B
H1 : Variable A est dépendante de la variable 2
Chi2 =15,1, P value < 0,01



R=0,91 - P value <0,05

HO : R est égal à 0
H1 : R est significativement différent de O

	blonds	bruns	noirs	roux	Total
bleus	25	9	3	7	44
gris	13	17	10	7	47
marrons	7	13	8	5	33
Total	45	39	21	19	124

Répartition couleur cheveux/yeux

Chi2 =15,1, P value < 0,01

H0 : la couleur des yeux est indépendante de la couleur des cheveux
H1 : H0 : la couleur des yeux est dépendante de la couleur des cheveux

https://statistics.ead-minerve.fr

Matrice de corrélations

http://www.sthda.com/french/wiki/matrice-de-correlation-avec-r-analyse-et-visualisation

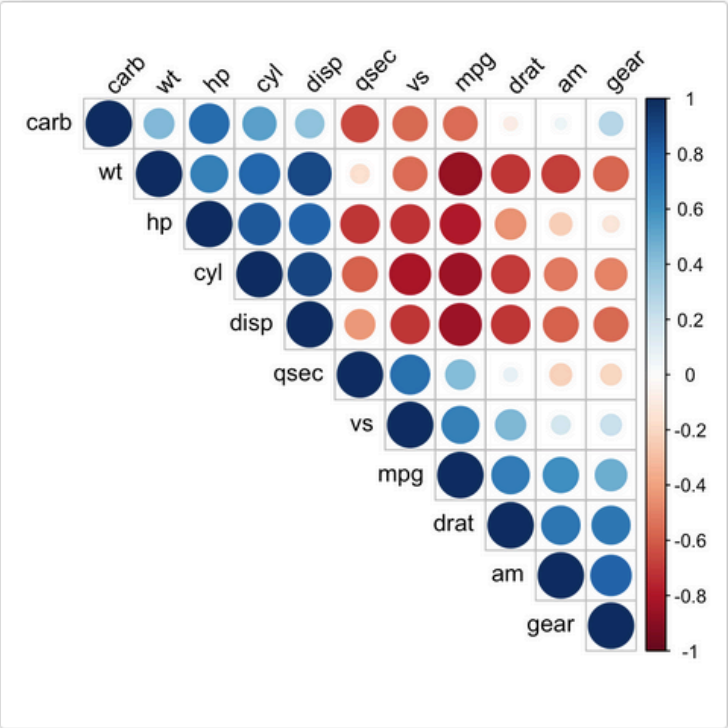
```r
library(Hmisc)
rcorr(as.matrix(mtcars[,1:7]))
```

```
       mpg   cyl  disp    hp  drat    wt  qsec
mpg   1.00 -0.85 -0.85 -0.78  0.68 -0.87  0.42
cyl  -0.85  1.00  0.90  0.83 -0.70  0.78 -0.59
disp -0.85  0.90  1.00  0.79 -0.71  0.89 -0.43
hp   -0.78  0.83  0.79  1.00 -0.45  0.66 -0.71
drat  0.68 -0.70 -0.71 -0.45  1.00 -0.71  0.09
wt   -0.87  0.78  0.89  0.66 -0.71  1.00 -0.17
qsec  0.42 -0.59 -0.43 -0.71  0.09 -0.17  1.00
n= 32
P
     mpg    cyl    disp   hp     drat   wt     qsec
mpg         0.0000 0.0000 0.0000 0.0000 0.0000 0.0171
cyl  0.0000        0.0000 0.0000 0.0000 0.0000 0.0004
disp 0.0000 0.0000        0.0000 0.0000 0.0000 0.0131
hp   0.0000 0.0000 0.0000        0.0100 0.0000 0.0000
drat 0.0000 0.0000 0.0000 0.0100        0.0000 0.6196
wt   0.0000 0.0000 0.0000 0.0000 0.0000        0.3389
qsec 0.0171 0.0004 0.0131 0.0000 0.6196 0.3389
```

Matrice de corrélations

http://www.sthda.com/french/wiki/matrice-de-correlation-avec-r-analyse-et-visualisation

```
library(corrplot)
corrplot(mcor, type="upper", order="hclust", tl.col="black", tl.srt=45)
```



Les **corrélations positives** sont affichées en bleu et les **corrélations négatives** en rouge. L'intensité de la couleur et la taille des cercles sont proportionnelles aux **coefficients de corrélation**. A droite du **corrélogramme**, la légende de couleurs montre les **coefficients de corrélation** et les couleurs correspondantes.

Tableaux des fréquences

fréquence	blond	bruns	noir	roux	Total
bleu	25	9	3	7	44
gris	13	17	10	7	47
marron	7	13	8	5	33
Total	45	39	21	19	124

%total	blond	bruns	noir	roux	Total
bleu	20,2	7,3	2,4	5,6	35,5
gris	10,5	13,7	8,1	5,6	37,9
marron	5,6	10,5	6,5	4,0	26,6
Total	36,3	31,5	16,9	15,3	100,0

%ligne	**blond**	**bruns**	**noir**	**roux**	**Total**
bleu	56,8	20,5	6,8	15,9	100,0
gris	27,7	36,2	21,3	14,9	100,0
marron	21,2	39,4	24,2	15,2	100,0
Total	36,3	31,5	16,9	15,3	100,0

- Référence
  - Howell DC : méthodes statistiques en sciences humaines
    - De l’échantillonnage
    - Visualisations
    - Corrélation
    - Dépendance
    - Tests /anova
    - Etc …