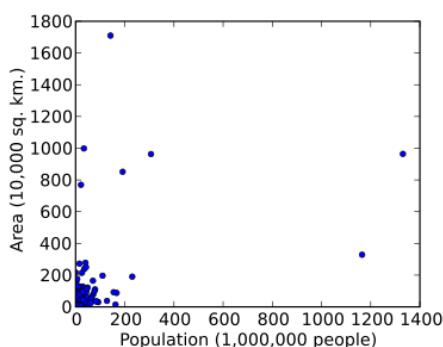


## 01. Fiche d'auto-évaluation

February 6, 2023

- Regardez le graphe ci-dessous<sup>1</sup>



Pouvez-vous voir clairement s'il y a une dépendance entre "Population" et "Area"? Y a-t-il des transformations qui peuvent aider à montrer plus clairement la possible dépendance? Si oui, est-il mieux d'appliquer telle transformation à l'axe des x, des y, ou aux deux?

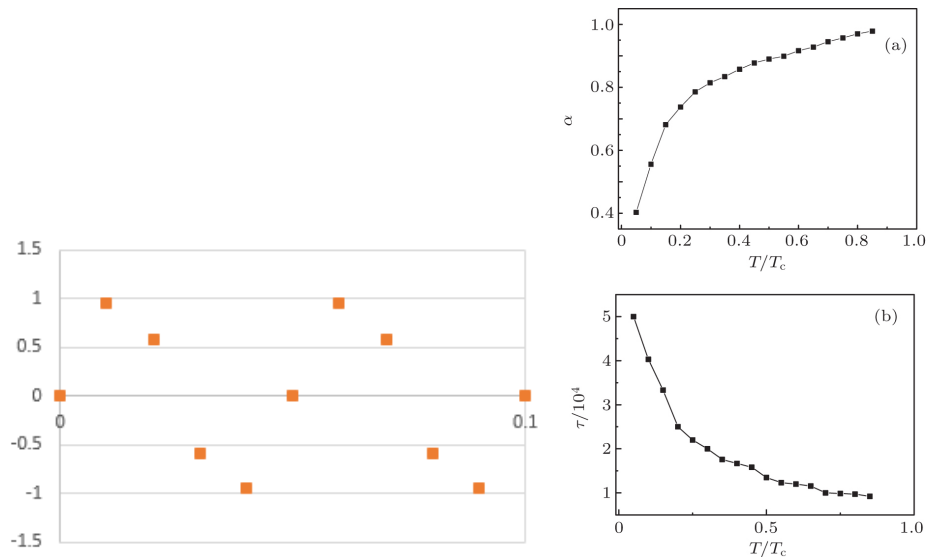
- Regardez les trois graphes ci-dessous <sup>2</sup> <sup>3</sup>

---

<sup>1</sup>Attribution Skbkekas, CC BY-SA 3.0 <<http://creativecommons.org/licenses/by-sa/3.0/>>, via Wikimedia Commons

<sup>2</sup>From [http://cpb.iphys.ac.cn/article/2015/cpb\\_24\\_7\\_070501/cpb142814f4\\_hr.jpg.html](http://cpb.iphys.ac.cn/article/2015/cpb_24_7_070501/cpb142814f4_hr.jpg.html)

<sup>3</sup>From P. Scheidler "Understanding the Basics of Fourier Transforms"



Pour chaque graphe, dites s'il y a une dépendance entre les x et les y. Si on calcule le coefficient de corrélation de Pearson, sera-t-il positif négatif ou zéro?

- Imaginez de travailler avec un jeu de donnée, dont une caractéristique peut prendre 4 valeurs. Vous avez deux façons de l'encoder: (i) transformer les valeurs en 0,1,2,3 ou (ii) appliquer le one-hot encoding. Dans quel cas faut-il choisir l'un ou l'autre?
- Imaginez de travailler avec un jeu de donnée de 200 colonnes. Une colonne a 0.1% de valeurs manquantes. Une autre colonne a 30% de valeurs manquantes. Quel type de traitement (pre-processing) peut-on appliquer au jeu de données avant d'appliquer d'autres méthodes supervisées ou non-supervisées? [Regardez le code du notebook concernant le pre-processing] Imaginez qu'une de vos colonnes aie les valeurs suivantes: étudiant, médecin, ingénieur, cassier, info-non-disponible. Comment peut-on transformer la colonnes pour permettre aux méthodes d'apprentissage automatique de l'utiliser?
- Imaginez qu'une de vos colonnes aie les valeurs suivantes: very bad, bad, medium, good, very good, excellent, no-opinion. Comment peut-on transformer la colonnes pour permettre aux méthodes d'apprentissage automatique de l'utiliser?
- Écrivez les formules de la variance d'une caractéristique, l'écart type (standard deviation), le coefficient de corrélation de Pearson.
- Expliquez comment on construit une boîte à moustache (boxplot) et la signification de toutes les lignes qu'on y retrouve.
- Expliquez comment on calcule le 35me percentile.

