# Visualisation de données: populations microbiennes issues des toilettes publiques

Marie Martelat et Johanna Galvis. M2 Bio-informatique, UCBL Lyon 1.

#### 1. Introduction

En utilisant les données issus de la publication "Microbial Biogeography of Public Restroom Surfaces" de Flores et al., nous voulons montrer les possibilités qu'offfent les packages Python pour la visualization de données dans le contexte de la métagénomique.

Nous avions à notre disposition une table de comptage de 14 échantillons.

Les échantillons ont été collectés sur différentes surface de toilette. Surface touché par les mains (door\_in, faucet\_handle, toilet\_flush\_handle) par les pieds (sink\_floor, toilet\_floor) ou en contacte directe avec la partie posterieur du corp(toilet\_seat).

Il peut être intéressant d'essayer de visualiser une différence de composition bactérienne dans ces différents type d'echantillons.

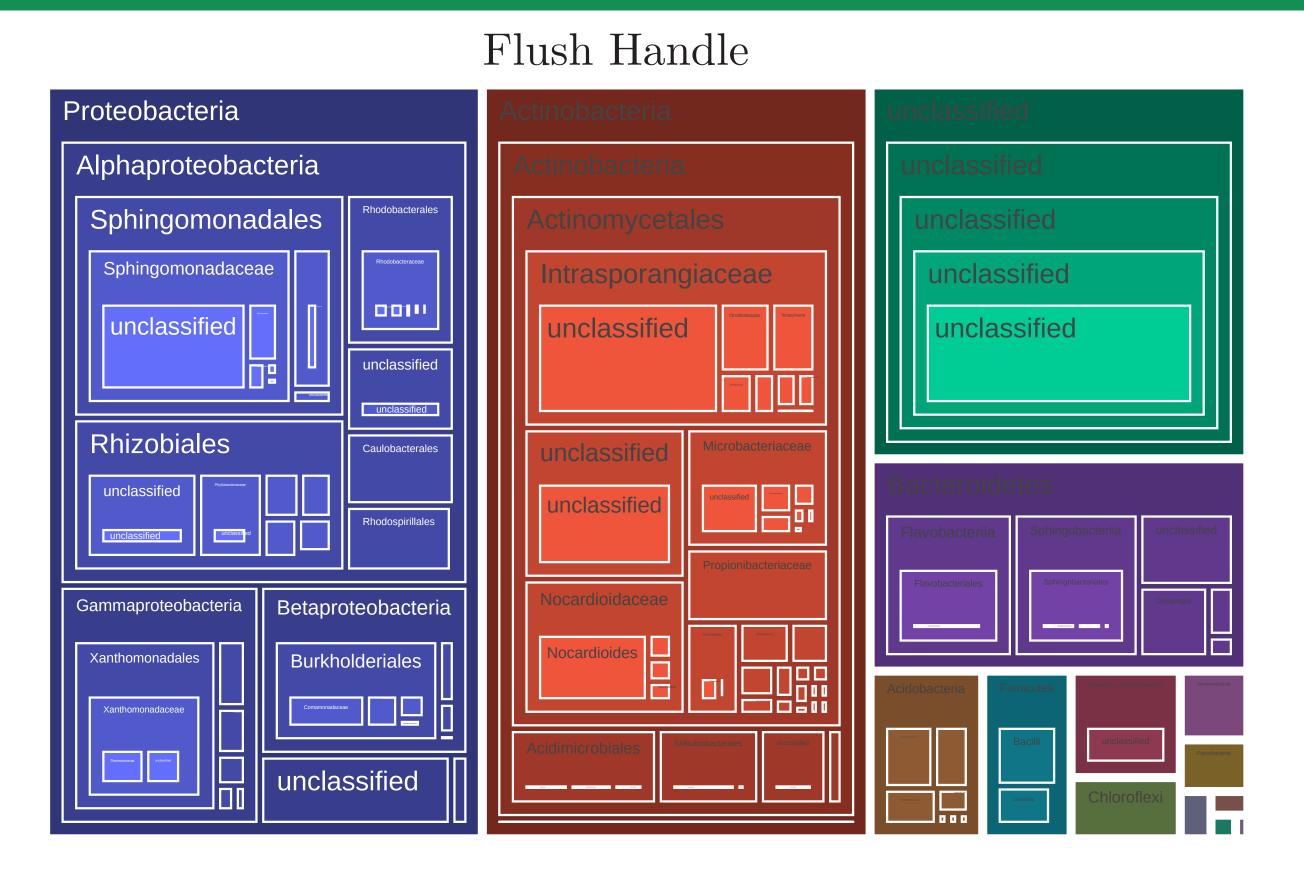
#### 2.BarPlot Abondance relative des phylums, sur différentes surfaces. Quatre taxons ressortent quelque soit la surface (Actinobacteria, Proteobacteria, fimicutes, Bacteroidetes). Abondances relative (Niveau : Phylum ; cutoff counts =250) Actinobacteria(100) toilet seat 2 Proteobacteria(100) toilet\_seat\_1 Firmicutes(100) toilet flush handle 2 Bacteroidetes(100) toilet flush handle 1 unclassified(100) toilet floor 2 toilet floor 1 Parcubacteria(100) stall\_in\_1 soap\_dispenser\_1 sink floor 2 sink floor 1 faucet handle 2 faucet handle 1 door\_in\_2 door\_in\_1 3000 500 2000 2500 1000 1500 Abondance

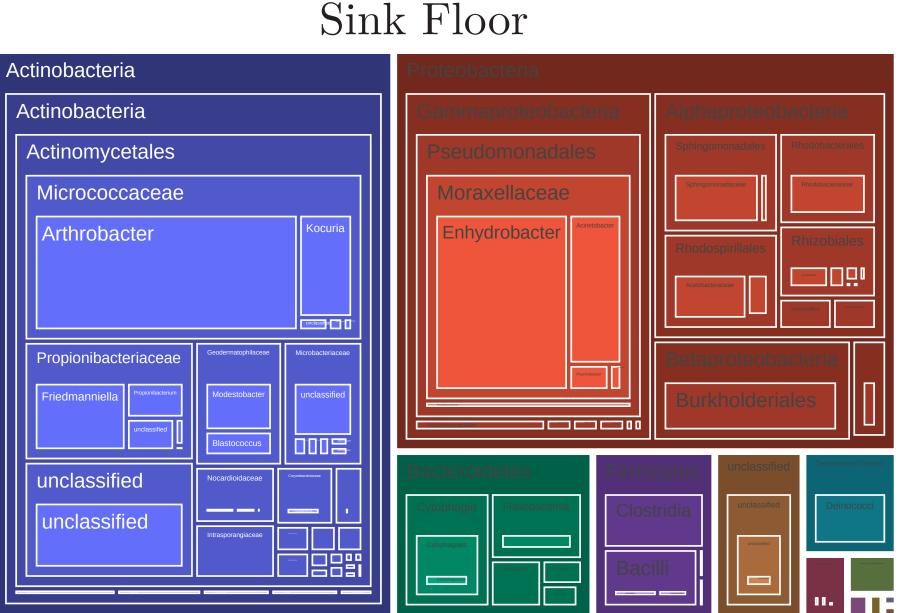
#### 3.HeatMap Visualisation de l'abscence (vert) ou présence (jaune) de taxon sur les différentes surfacess (Niveaux Phylum). Présence/Abscence d'un Phylum door in 1 door in 2 faucet handle 1 faucet\_handle\_2 sink floor 1 sink floor 2 soap\_dispenser\_1 stall in 1 toilet floor toilet floor 2 toilet\_flush\_handle\_1 toilet flush handle 2 toilet seat 1 toilet\_seat\_2

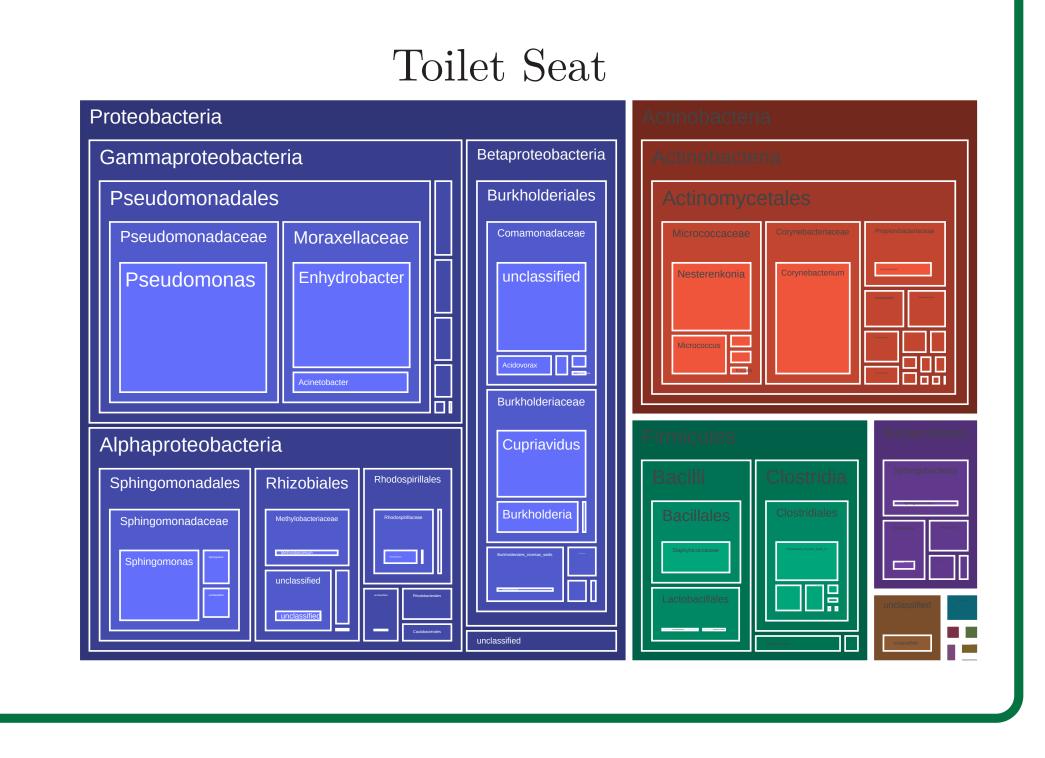
### 5.TreeMap

Treemap prenant en compte la hiérachie des taxons.

On pars du Regne ( Un seul réprésenté : Bacteria), puis le Phylum (2 groupes ressortent: Actinobacteria, Proteobacteria), ensuite la classe, l'ordre, la famille et enfin le genre.





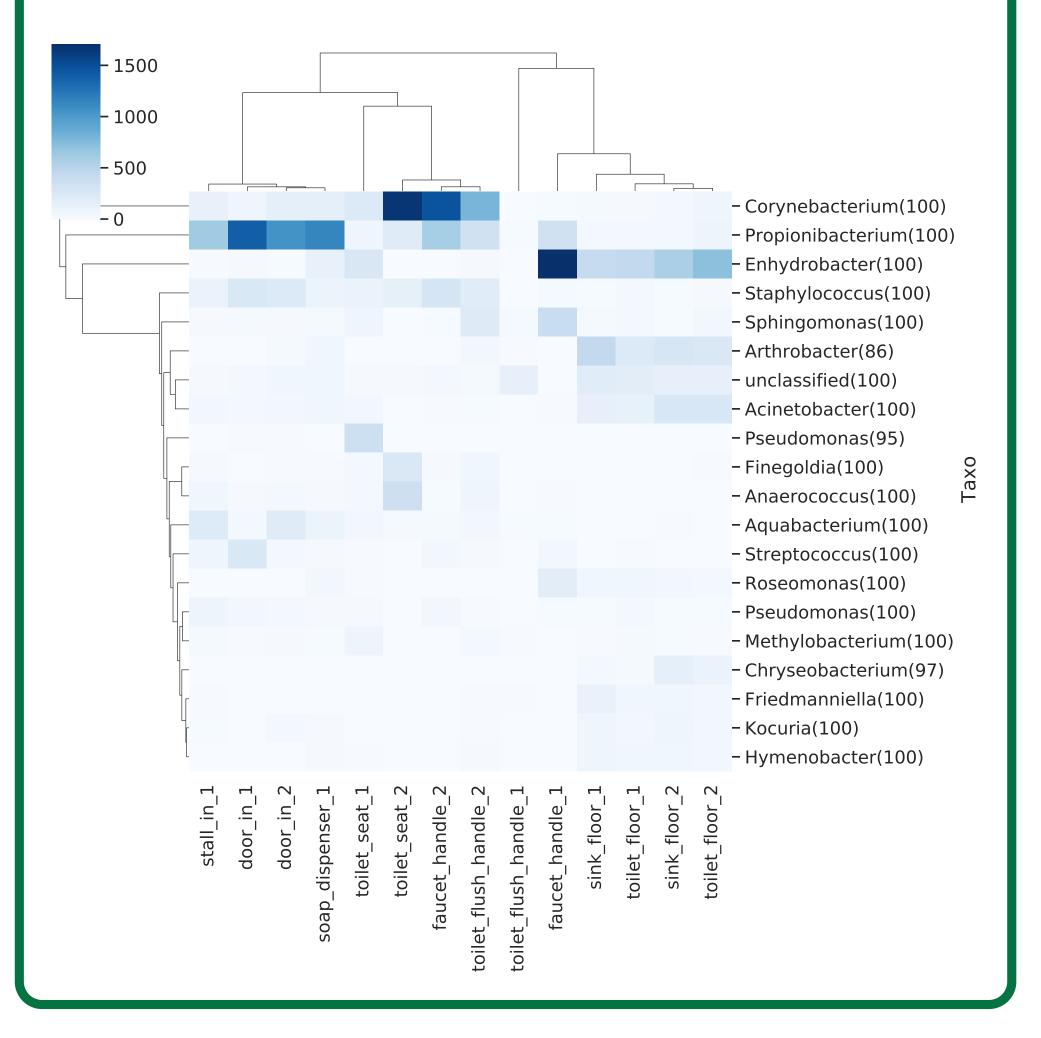


# Conclusion

Des puissants outils de visualisation permettent la découverte d'information dans les données. Malgré les limitations propres au librairies utilisées, des *patterns* intéressants ont pu être identifiés. Des particularités des échantillons doivent être tenues en compte (une composition innatendue).

# 4. Clustering Hierarchique

Au niveau genre, le calcul des distances entre echantillons laisse aprécier des groupements selon la surface corporelle impliquée.



## 6. ACP

Notez la proximité innatendue entre échantillons du sols (bleue), toilet flush handle(vert) et robinet(citron).

