

Introduction à R

Exercices sur les données des mairies

Pascal Bessonneau

06/2015

La sécurité informatique dans les mairies

Source

Ces données proviennent du site data.gouv.fr.

Ce sont des données sur la sécurisation des sites web des mairies.

Chargement

```
> mairies <- read.csv("data/ville_secu_informatique_centroide.csv"  
+                      , fileEncoding = "UTF8"  
+                      , stringsAsFactors = F  
+                      )
```

Résumé

```
> summary(mairies)
```

```
##      Commune           Code.Insee
## Length:14220      Length:14220
## Class :character   Class :character
## Mode  :character   Mode  :character
##
##
##
##      url           Catégorie
## Length:14220      Length:14220
## Class :character   Class :character
## Mode  :character   Mode  :character
##
##
##
##      Population           Site
## Min.   :      19      Length:14220
## 1st Qu.:     565      Class :character
## Median :    1176      Mode  :character
## Mean   :    4123
```

Résumé

```
> prop.table(table(mairies$Site))

##
## oui
##    1

> prop.table(table(mairies$Site))*100

##
## oui
## 100
```

Questions

- Quelle est la proportion de mairies avec un accès HTTPS ?
- Quelle est la proportion de chaque catégorie ?
- Quelle est la proportion de chaque type de serveur ?
- Faire le tableau croisé HTTPS/Serveur.
- Faire le tableau croisé Serveur/catégorie.
- Est-ce que la taille moyenne de la population varie selon le type de serveur ?
- Est-ce que la taille moyenne de la population varie en fonction du type de langage ?
- Est-ce que la taille moyenne de la population varie en fonction du type de langage ?

Questions

Quelle est la proportion de mairies avec un accès HTTPS ?

```
> prop.table(table(mairies$https))*100
```

```
##
```

```
##      non      oui
```

```
## 98.924051 1.075949
```


Questions

Quelle est la proportion de chaque catégorie ?

```
> prop.table(table(mairies$Catégorie))*100
```

```
##
##                               À jour
##                               7.144866
## Informations masquées (vulnérabilités pas exposées)
##                               27.208158
##                               Niveau de sécurité incertain
##                               19.099859
##                               Pas à jour
##                               46.547117
```

Questions

Quelle est la proportion de chaque catégorie ?

```
> prop.table(table(mairies$Catégorie))*100
```

```
##
##                               À jour
##                               7.144866
## Informations masquées (vulnérabilités pas exposées)
##                               27.208158
##                               Niveau de sécurité incertain
##                               19.099859
##                               Pas à jour
##                               46.547117
```

Questions

Quelle est la proportion de chaque type de serveur ?

```
> prop.table(table(mairies$Serveur))*100
```

##			
##	Apache	Apache Tomcat	Google
##	76.49085795	0.16174402	0.42897328
##	Icodia	IIS	Inconnue
##	0.08438819	8.07313643	1.61040788
##	Lighttp	LiteSpeed	Lotus-Domino
##	1.34317862	0.26722925	0.04922644
##	Nginx	o2switch	SiteW
##	10.35161744	0.30942335	0.33052039
##	Tengine	Wix	WMaker
##	0.06329114	0.26722925	0.16877637

Questions

Quelle est la proportion de chaque type de serveur ?

```
> prop.table(table(mairies$Serveur,mairies$https))*100
```

```
##
##              non              oui
##  Apache      75.787623066  0.703234880
##  Apache Tomcat 0.154711674  0.007032349
##  Google      0.246132208  0.182841069
##  Icodia      0.084388186  0.000000000
##  IIS         8.073136428  0.000000000
##  Inconnue    1.547116737  0.063291139
##  Lighthttp   1.343178622  0.000000000
##  LiteSpeed   0.267229255  0.000000000
##  Lotus-Domino 0.049226442  0.000000000
##  Nginx      10.288326301  0.063291139
##  o2switch    0.309423347  0.000000000
##  SiteW      0.323488045  0.007032349
##  Tengine    0.063291139  0.000000000
##  Wix        0.218002813  0.049226442
##  WMaker     0.168776371  0.000000000
```

Questions

Quelle est la proportion de chaque type de serveur ?

```
> prop.table(table(mairies$Serveur,mairies$Catégorie))*100
```

```
##
##           À jour
##  Apache      6.772151899
##  Apache Tomcat 0.000000000
##  Google      0.000000000
##  Icodia      0.000000000
##  IIS         0.021097046
##  Inconnue    0.028129395
##  Lighttpd    0.000000000
##  LiteSpeed   0.112517581
##  Lotus-Domino 0.000000000
##  Nginx       0.154711674
##  o2switch    0.035161744
##  SiteW       0.021097046
##  Tengine     0.000000000
##  Wix         0.000000000
##  WMaker      0.000000000
##
```

Questions

Est-ce que la taille moyenne de la population varie selon le type de serveur ?

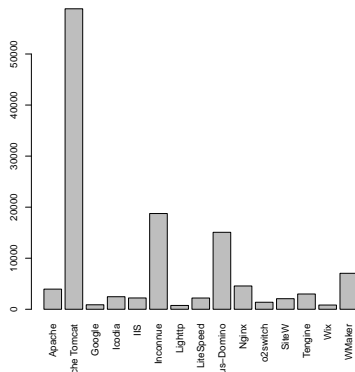
```
> (tmp <- tapply(mairies$Population,mairies$Serveur,mean))
```

##	Apache	Apache Tomcat	Google
##	3945.7721	58853.7826	879.1803
##	Icodia	IIS	Inconnue
##	2457.6667	2213.8841	18754.1048
##	Lighttp	LiteSpeed	Lotus-Domino
##	746.8482	2203.5526	15075.0000
##	Nginx	o2switch	SiteW
##	4563.7548	1374.8409	2070.4468
##	Tengine	Wix	WMaker
##	2991.6667	828.5263	7045.2917

Questions

Est-ce que la taille moyenne de la population varie selon le type de serveur ?

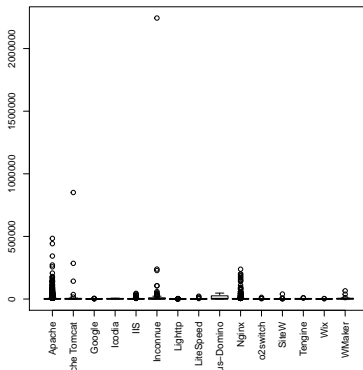
```
> barplot(tmp, las=3)
```



Questions

Est-ce que la taille moyenne de la population varie selon le type de serveur ?

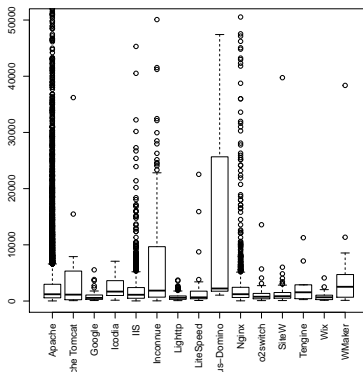
```
> boxplot(mairies$Population ~ mairies$Serveur, las=3)
```



Questions

Est-ce que la taille moyenne de la population varie selon le type de serveur ?

```
> boxplot(mairies$Population ~ mairies$Serveur, las=3, ylim=c(0,50000))
```



Création de la variable département

substr est une fonction qui permet de sélectionner une partie de la chaîne de caractère.

```
> mairies$dept <- substr(mairies$Code.Insee,1,2)
```

Création de la variable département

```
> prop.table(table(mairies$dept,mairies$Catégorie),1)*100
```

```
##  
##      À jour  
## 01 15.319149  
## 02  4.511278  
## 03  3.225806  
## 04  5.405405  
## 05  7.462687  
## 06  3.278689  
## 07 43.030303  
## 08  4.545455  
## 09  6.060606  
## 10  1.298701  
## 11  5.072464  
## 12  1.538462  
## 13  6.837607  
## 14  6.521739  
## 15  2.531646  
## 16  2.678571  
## 17  4.000000
```