Analyse réplicable d'un jeu de données

Christophe Genevey

13 septembre 2017

Introduction

L'objectif de cet exercice est d'analyser un jeu de données. Ce jeu de données est disponible sur ce lien

Présentation des données

```
NUMID : Matricule anonyme de la personne interrogée

CONNAITRE : Réponse à la question "Connaissez-vous les produits biologiques?"
  - 0 : pas de réponse
  - 1 : oui
  - 2 : non

DIFF : Réponse à la question "Y a t-il une différence entre le produit biologique et produit diététique
  - 0 : pas de réponse
  - 1 : oui
  - 2 : non

CONSOM : Réponse à la question "Avez-vous déjà consommé des produits biologiques?"
  - 1 : non jamais
  - 2 : oui une seule fois
  - 3 : oui rarement
  - 4 : oui de temps en temps
  - 5 : oui plusieurs fois par mois
  - 6 : oui plusieurs fois par semaine
  - 7 : pas de réponse

MARQUE : Réponse à la question "Parmis les marques suivantes lesquelles connaissez-vous?"
  - 0 : pas de réponse
  - 1 : bio vivre
  - 2 : bjorg
  - 3 : carrefour bio
  - 4 : la vie
  - 5 : vrai
  - 6 : prosain
  - 7 : favrichon

CONSVIE: Réponse à la question "Avez-vous déjà consommé des produits "La Vie"?
  - 0 : pas de réponse
  - 1 : oui une fois
  - 2 : oui occasionnellement
  - 3 : oui réguiliérement
  - 4 : non jamais

SEXE: Sexe de la personne
  - 1 : homme
```

```
  - 2 : femme

AGE: Age de la personne
  - 1 : moins de 25 ans
  - 2 : entre 25 et 35 ans
  - 3 : entre 35 et 45 ans
  - 4 : entre 45 et 55 ans
  - 5 : entre 55 et 65 ans
  - 6 : plus de 65 ans

ETATCIVIL: Etat-civil de la personne
  - 0 : autre
  - 1 : marie
  - 2 : celibataire
  - 3 : divoré
  - 4 : en concubinage
  - 5 : veuf

NBENF: Nombre d'enfants de la personne
  - 1 : 0 enfant
  - 2 : 1 enfant
  - 3 : 2 enfants
  - 4 : 3 enfants
  - 5 : plus de 3 enfants

SITPROF: Situation professionnelle
  - 1 : Agriculteur
  - 2 : Artisan
  - 3 : Cadre supérieur
  - 4 : Cadre moyen
  - 5 : Employé
  - 6 : Ouvrier
  - 7 : Retraité
  - 8 : Autre
  - 9 : pas de réponse

REVENUE: Revenus mensuels de la personne
  - 0 : pas de réponse
  - 1 : moins de 5 kF
  - 2 : entre 5 et 10 kF
  - 3 : entre 10 et 15 kF
  - 4 : entre 15 et 20 kF
  - 5 : plus de 20 kF
  - 6 : ne se prononce pas
```

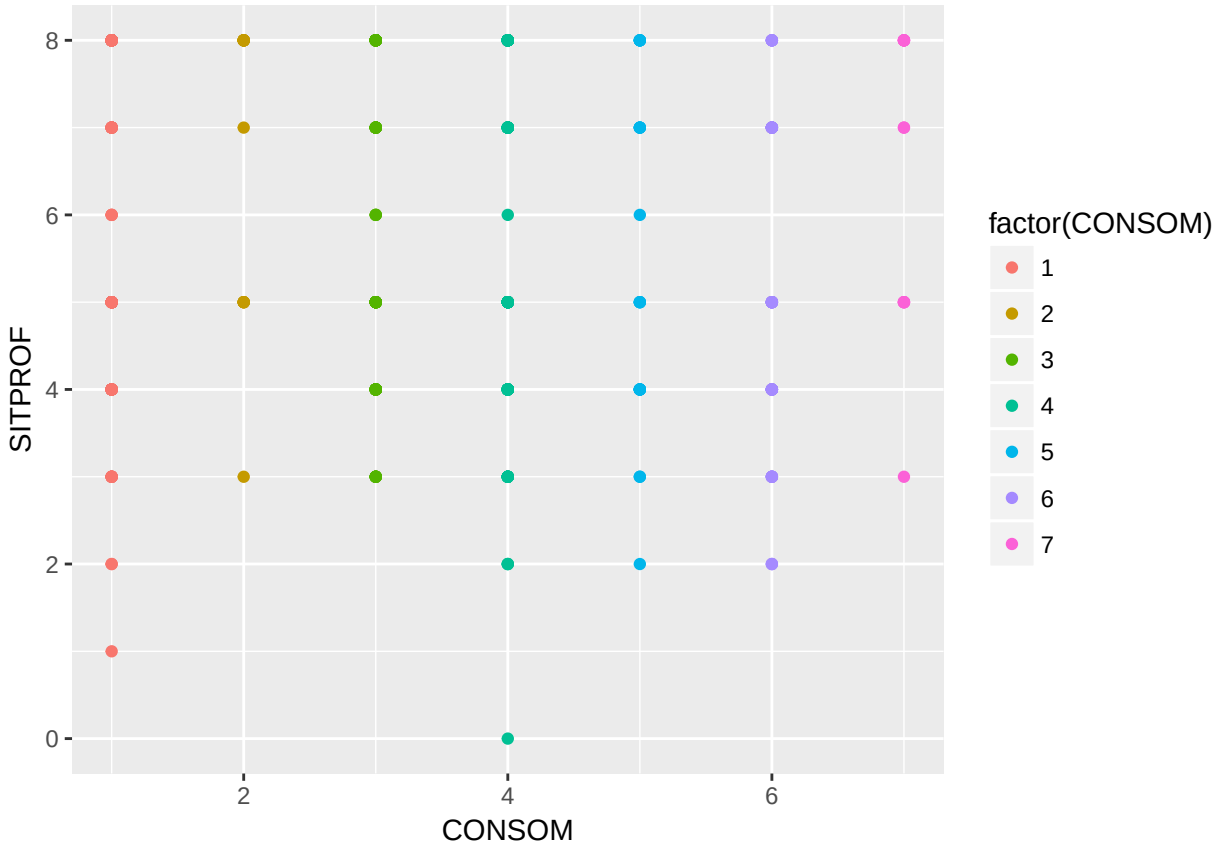
Jeu de données

```
library(ggplot2)
dataset <- read.csv(file="data/pbio.csv", header=TRUE)
summary(dataset)
```

```
##       CODE         CONNAITRE          DIFF           CONSOM
```

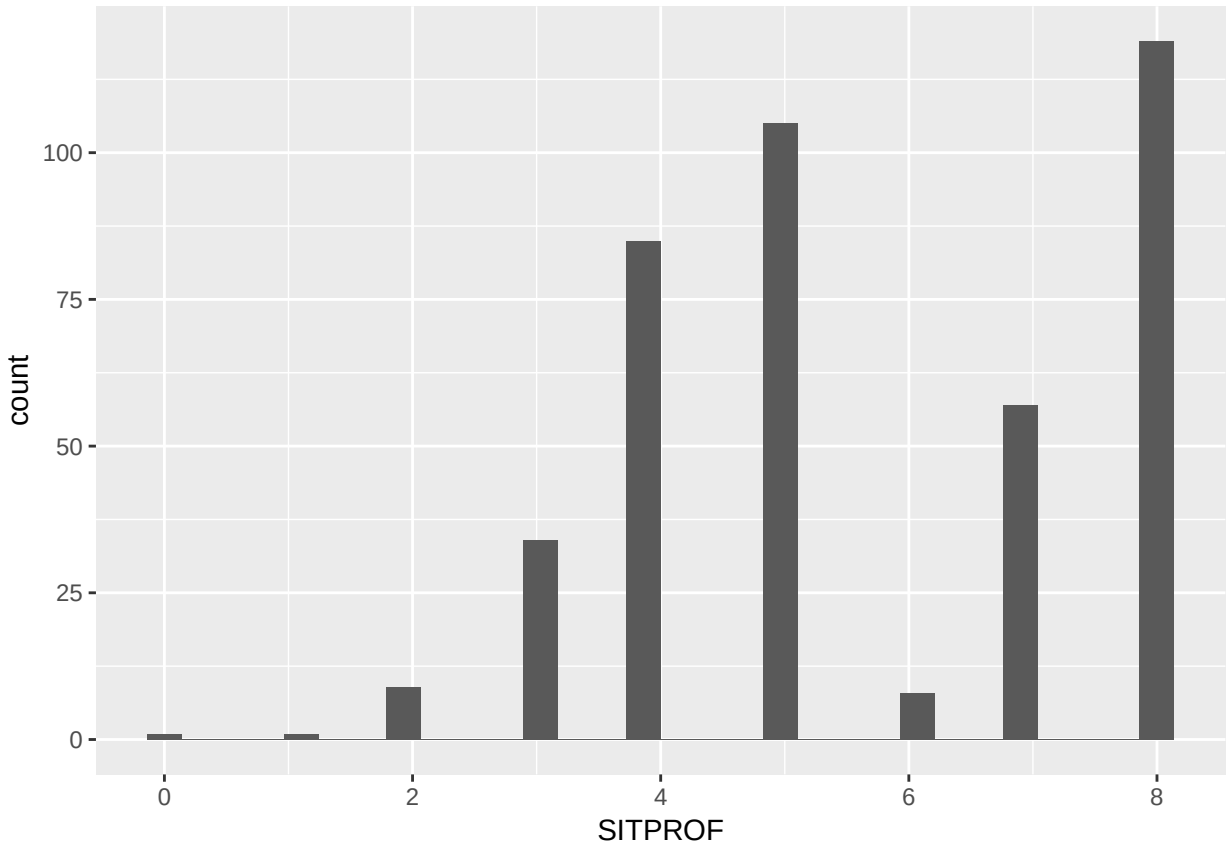
```
##  Min.   :10101   Min.   :0.000   Min.   :0.000   Min.   :1.000
##  1st Qu.:11102   1st Qu.:1.000   1st Qu.:1.000   1st Qu.:1.000
##  Median :30405   Median :1.000   Median :1.000   Median :3.000
##  Mean   :25654   Mean   :1.031   Mean   :1.203   Mean   :3.313
##  3rd Qu.:40402   3rd Qu.:1.000   3rd Qu.:1.000   3rd Qu.:4.000
##  Max.   :41512   Max.   :2.000   Max.   :2.000   Max.   :7.000
##      MARQUE          CONSVIE          SEXE            AGE
##  Min.   :0.000   Min.   :0.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:2.000   1st Qu.:4.000   1st Qu.:1.000   1st Qu.:2.000
##  Median :2.000   Median :4.000   Median :2.000   Median :3.000
##  Mean   :2.685   Mean   :3.594   Mean   :1.656   Mean   :3.191
##  3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:2.000   3rd Qu.:4.000
##  Max.   :7.000   Max.   :4.000   Max.   :2.000   Max.   :6.000
##    ETATCIVIL         NBENF          SITPROF         REVENU
##  Min.   :0.000   Min.   :1.000   Min.   :0.000   Min.   :0.000
##  1st Qu.:1.000   1st Qu.:1.000   1st Qu.:4.000   1st Qu.:2.000
##  Median :1.000   Median :1.000   Median :5.000   Median :4.000
##  Mean   :1.747   Mean   :1.795   Mean   :5.692   Mean   :3.628
##  3rd Qu.:2.000   3rd Qu.:3.000   3rd Qu.:8.000   3rd Qu.:5.000
##  Max.   :5.000   Max.   :5.000   Max.   :8.000   Max.   :6.000
```

```
ggplot(dataset, aes(x=CONSOM, y=SITPROF, color=factor(CONSOM))) + geom_point()
```



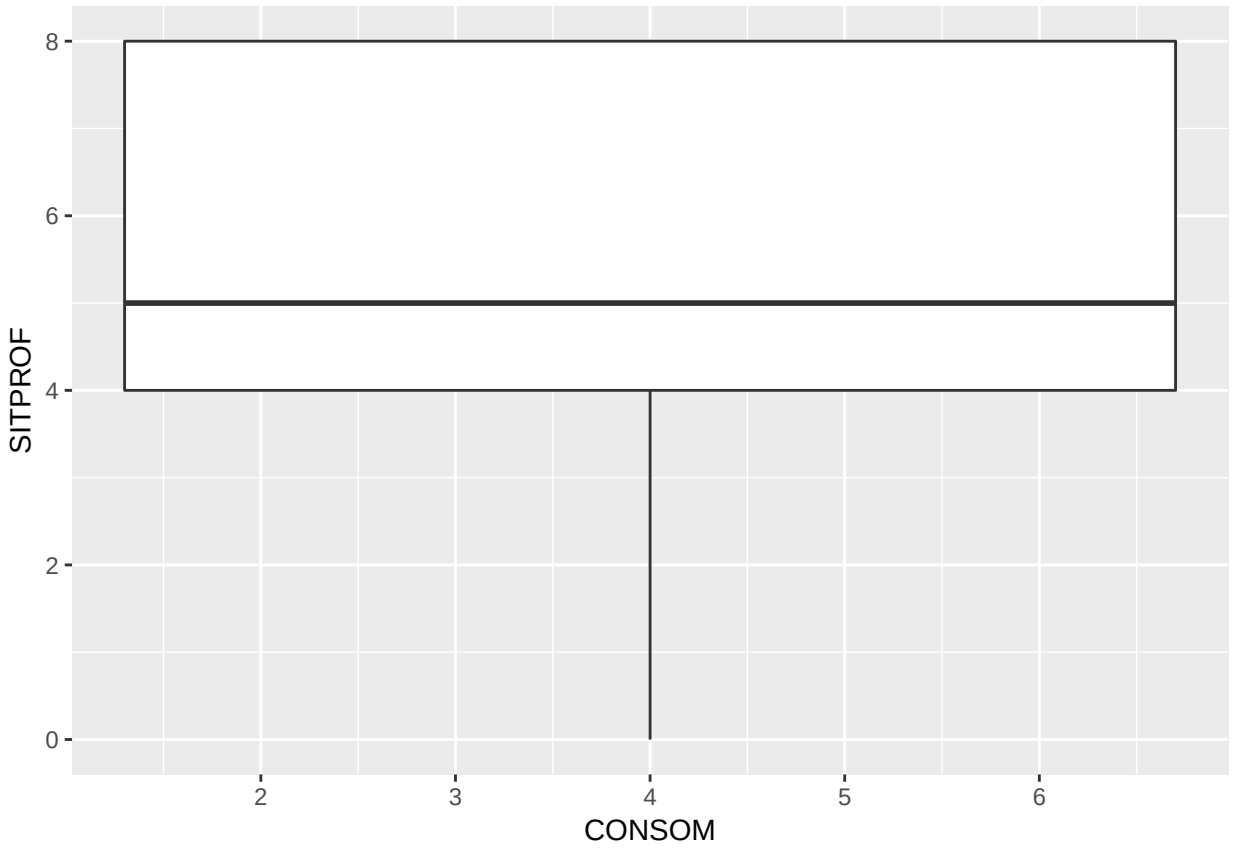

```
ggplot(dataset, aes(x=SITPROF)) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
ggplot(dataset, aes(x=CONSOM, y=SITPROF, color=CONSOM)) + geom_boxplot()
```

```
## Warning: Continuous x aesthetic -- did you forget aes(group=...)?
```


8
6
4
2
0
SITPROF
2
3
4
5
6
CONSOM