

Analyse répliquable 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

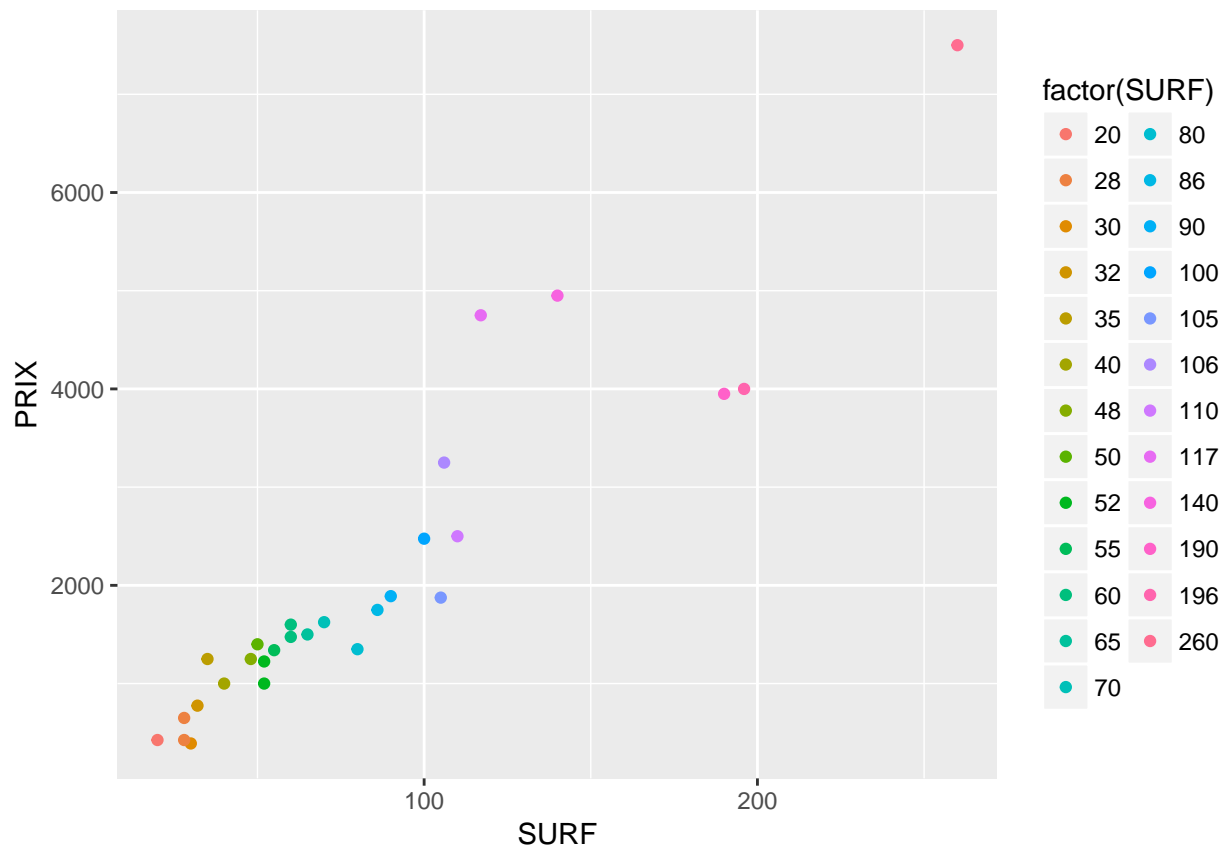
APP : Identifiant de l'appartement
SURF : Surface de l'appartement
PRIX : Prix de l'appartement
LOCALISATION : Localisation de l'appartement

Jeu de données

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

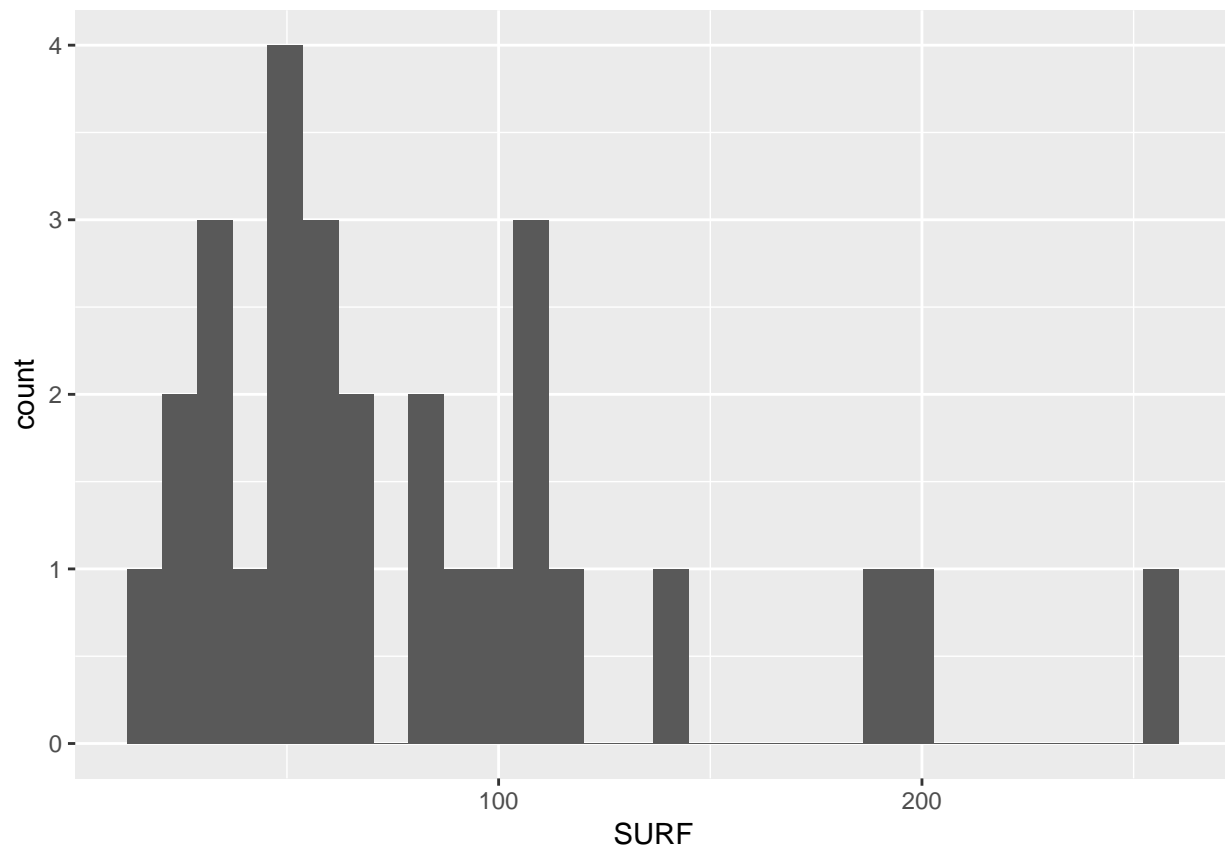
##	APP	SURF	PRIX	LOCALISATI
##	a01 : 1	Min. : 20.00	Min. : 390	assas luxembourg : 1
##	a02 : 1	1st Qu.: 46.00	1st Qu.:1169	boulevard saint ger: 1
##	a03 : 1	Median : 62.50	Median :1488	censier : 1
##	a04 : 1	Mean : 82.32	Mean :2056	censier 1 : 1
##	a05 : 1	3rd Qu.:105.25	3rd Qu.:2481	censier 2 : 1
##	a06 : 1	Max. :260.00	Max. :7500	contrescarpe : 1
##	(Other):22			(Other) :22

```
ggplot(dataset, aes(x=SURF, y=PRIX, color=factor(SURF))) + geom_point()
```



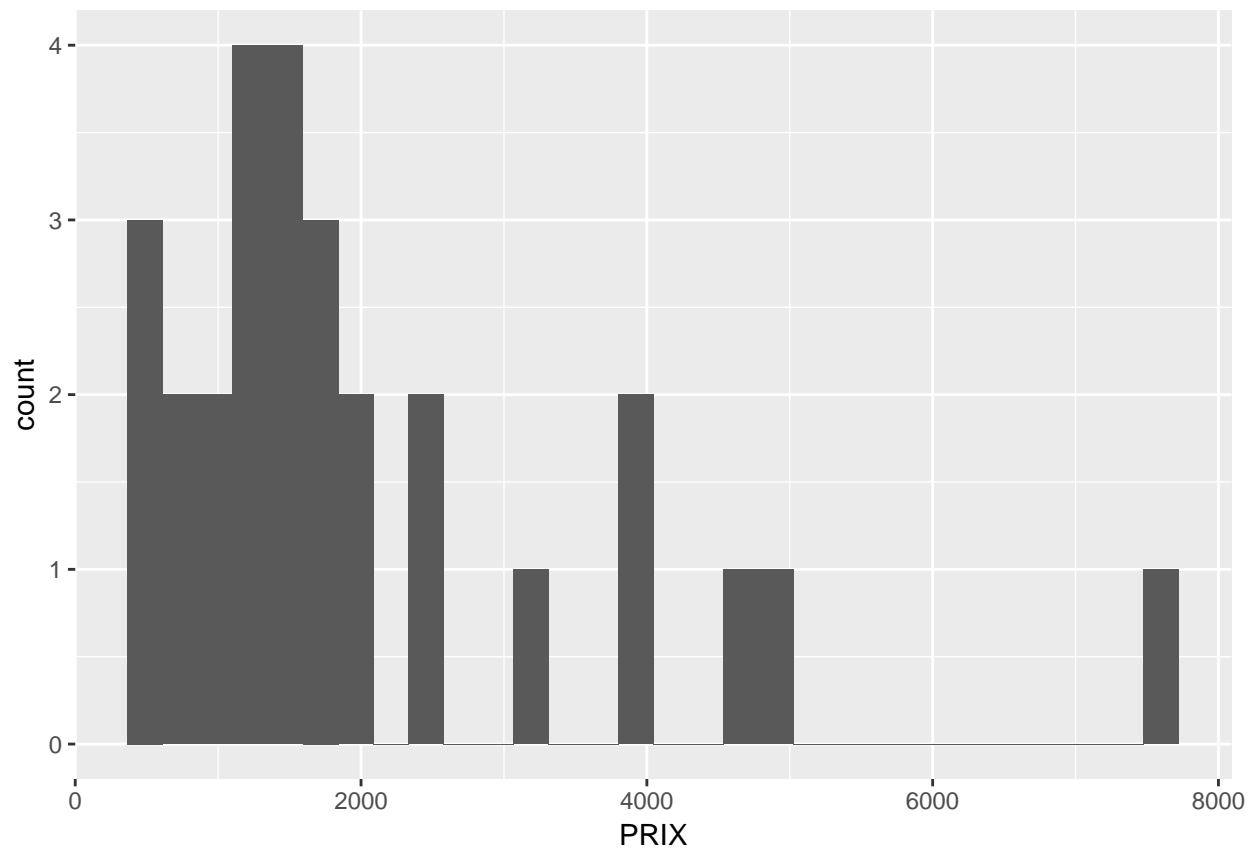
```
ggplot(dataset, aes(x=SURF)) + geom_histogram()
```

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



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

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



```
ggplot(dataset, aes(x=SURF, y=PRIX, color=factor(SURF))) + geom_boxplot()
```

