Un premier test en RMarkdown

Laurent PELLET

15/09/2016

Introduction

L'Étude Budget des Familles 2000 a été réalisée par l'Institut de la Statistique de la Polynésie Française (ISPF) entre mai 2000 et mai 2001 et a concerné plus de 3 600 ménages sur l'ensemble du terri- toire polynésien. Elle fait suite à une première opération similaire entreprise en 1986 mais de manière moins complète (1 062 ménages enquêtés dans les Îles du Vent) et exploitée de manière très partielle : seuls les élé- ments nécessaires à l'élaboration de l'indice des prix de détail à la consommation avaient été utilisés.

apply, sapply, lapply

```
# it's not really long and ugly
# it just squares the input
# but imagine that it is long and ugly

m <- matrix(data=cbind(rnorm(30, 0), rnorm(30, 2), rnorm(30, 5)), nrow=30, ncol=3)

apply(m, 1, mean)
```

```
##  [1] 2.4230289 3.1943178 3.6363494 3.2647832 2.5626176 2.4051293 1.5649667
##  [8] 1.5570095 1.7037580 3.4148656 2.6662604 2.1687845 2.5872108 0.3083157
## [15] 3.1786464 1.7450262 2.2864203 4.0918119 2.7955752 2.2492124 2.6048116
## [22] 2.0839951 1.8762179 2.8061470 1.8438639 3.0461532 2.0902953 3.4042808
## [29] 2.4501557 1.9869359
```

```
apply(m, 2, mean)
```

```
## [1] 0.06347216 2.13492628 5.20129619
```

```
apply(m, 2, function(x) length(x[x<0]))
```

```
## [1] 16  0  0
```

```
sapply(1:3, function(x) x^2)
```

```
## [1] 1 4 9
```

```
sapply(1:3, function(x) mean(m[,x]))
```

```
## [1] 0.06347216 2.13492628 5.20129619
```

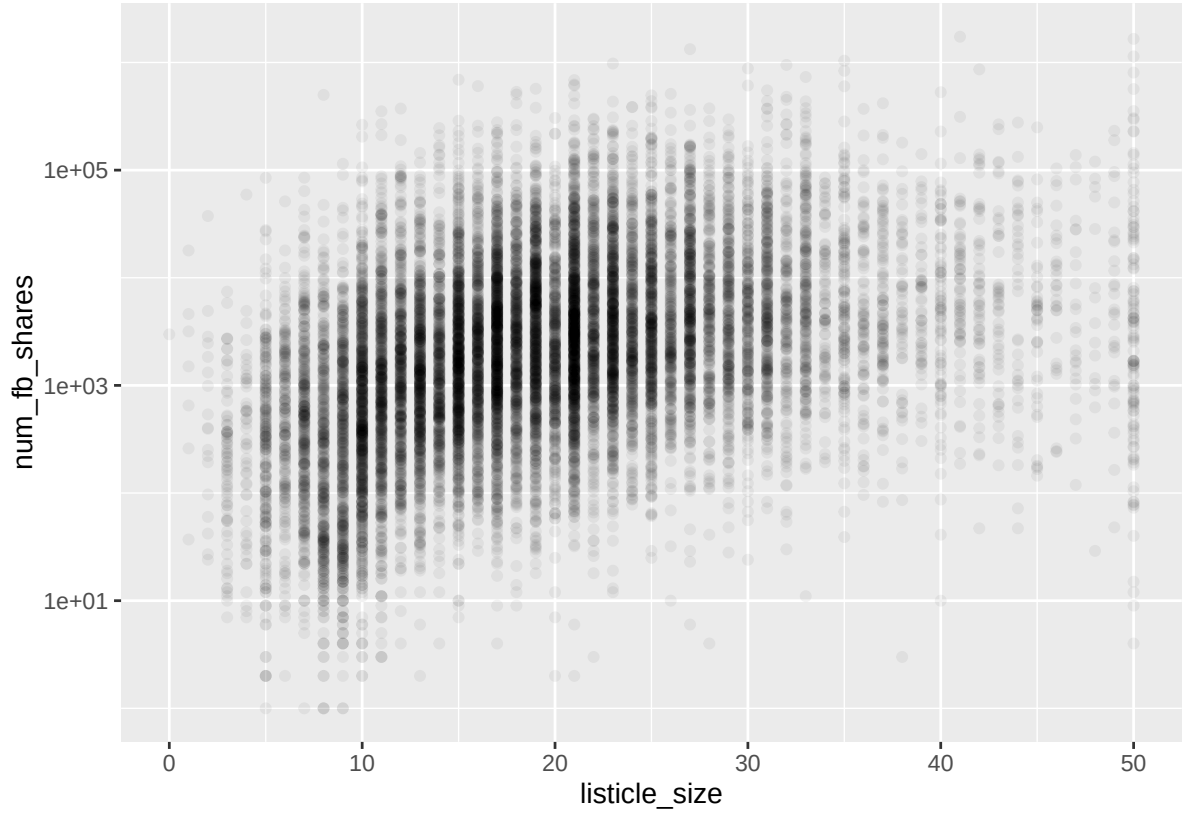
```
unlist(lapply(1:3, function(x) x^2))
```

```
## [1] 1 4 9
```

ggplot2

Because you are, it is not because i do there are a few listicles with over 1 million Facebook shares (welcome to 2015), the entire plot is skewed. As a result, we need to compress the plot by scaling the y-axis logarithmically using scale_y_log10. Additionally, there will be a large amount of overlap between points due to the large sample size, so we need to greatly reduce the opacity of the points. (I set to 5% for this chart, but the best value can be determined through trial and error)

```
library(ggplot2)
df <- read.csv("data/buzzfeed_linkbait_headlines.csv", header=T)
ggplot(df, aes(x=listicle_size, y=num_fb_shares)) +
  geom_point(alpha=0.05) +
  scale_y_log10()
```



Now we can apply the theme and labels