

Expression Data

Justine Guégan

7 mars 2017

Chargement et découverte du jeu de données

Lecture du jeu de données

```
data = read.table("gene1.txt", sep="\t", header=TRUE)
```

Description des groupes du jeu de données

```
table(data$group)
```

```
##
##      ctrl pre-sympto      sympto
##      20       30       50
```

Dimensions et aperçu du jeu de données

```
dim(data)
```

```
## [1] 100  2
```

```
knitr::kable(head(data))
```

	g1	group
4.155968		ctrl
4.489269		ctrl
4.796198		ctrl
3.622700		ctrl
7.312595		ctrl
5.285707		ctrl

Analyses

Quelques stats sur le jeu de données :

```
knitr::kable(summary(data))
```

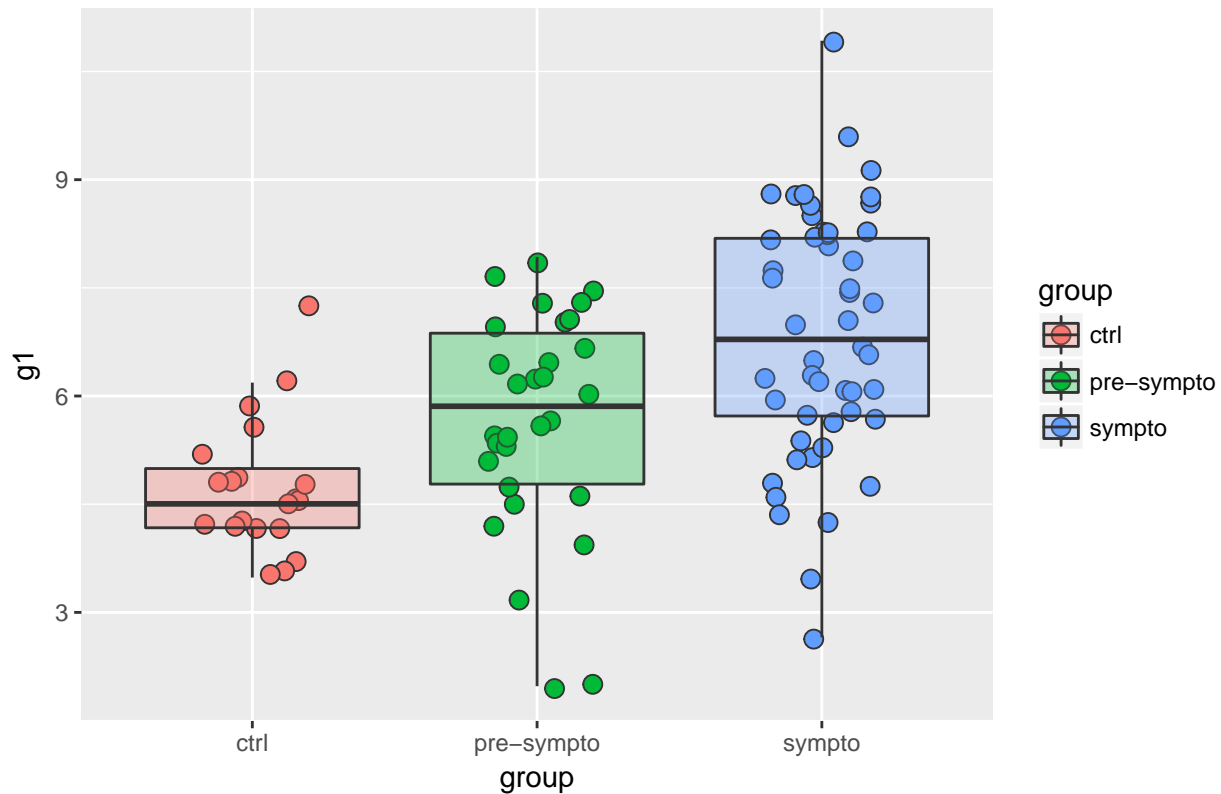
g1	group
Min. : 1.977	ctrl :20
1st Qu.: 4.796	pre-sympto:30
Median : 6.094	sympto :50
Mean : 6.083	NA
3rd Qu.: 7.396	NA
Max. :10.927	NA

Graphique

Visualisation pour le gene1 de l'expression par groupe :

```
library(ggplot2)
p2 = ggplot(data, aes(x=group, y=g1)) +
  geom_point(aes(fill=group), size=3, shape=21, colour="grey20",
    position=position_jitter(width=0.2, height=0.1)) +
  geom_boxplot(outlier.colour=NA, aes(fill=group), colour="grey20", alpha=0.3) +
  ggtitle("Gene 1 expression")
p2
```

Gene 1 expression



Statistique

T-test entre les contrôles et les pré-symptomatiques pour le gène1.

```
res1 = t.test(data$g1[which(data$group == "ctrl")], data$g1[which(data$group == "pre-sympto")])
res1
```

```
##
## Welch Two Sample t-test
##
## data: data$g1[which(data$group == "ctrl")] and data$g1[which(data$group == "pre-sympto")]
## t = -2.6624, df = 47.763, p-value = 0.01054
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -1.635714 -0.228028
## sample estimates:
## mean of x mean of y
## 4.737217 5.669088
```

Ajout de la pvalue sur le graphique précédent.

```
df1 <- data.frame(a = c(1,1,2,2), b = c(8.8, 9,9, 8.8))
p2 + geom_line(data = df1, aes(x = a, y = b)) +
  annotate("text", x = 1.5, y = 9.2,
    label = round(res1$p.value,2), size = 2)
```

Gene 1 expression

