Understanding RNA-seq analysis through an example: voting behavior at the French parliament

Hector Roux de Bézieux

May 2018

Introduction

Context

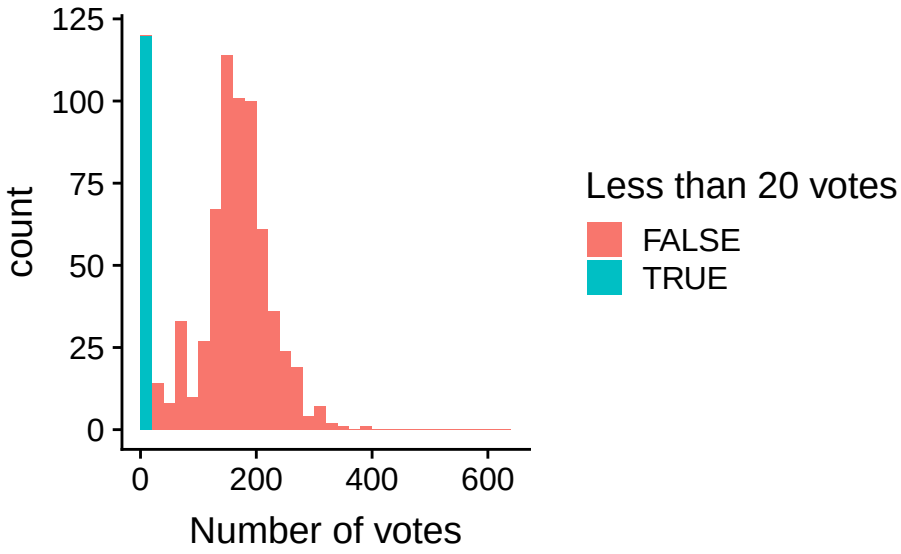
Loading the data and building an ExpressionSet object

We store the data in an *ExpressionSet* object, where the *phenodata* is the information about the cells (the deputees) and the *featuredata* is the information about the genes (the votes).

```
voting_record <- read_csv("data/voting_record.csv")
meta <- voting_record %>% select(circo, dept, name, surname, identifiant,
                                 iden, chamber, NbVote, NbYes, NbNo, NbAbst)
voting_record <- voting_record %>% select(-one_of(colnames(meta))) %>%
                                   t(.)
scrutins <- read_csv("data/scrutins.csv")
colnames(voting_record) <- rownames(meta) <- meta$identifiant
rownames(voting_record) <- rownames(scrutins) <- scrutins$Number
Assay <- ExpressionSet(assayData = voting_record,
                       phenoData = AnnotatedDataFrame(meta) ,
                       featureData = AnnotatedDataFrame(scrutins))
```

Filtering the cells (deputees)

```
Nb_Votes <- data.frame(Nb_Votes = colSums(!is.na(exprs(Assay))),
                       deputees = rownames(phenoData(Assay)))
ggplot(Nb_Votes, aes(x = Nb_Votes, fill = Nb_Votes <= 20)) +
                 stat_bin(breaks = seq(0, max(Nb_Votes$Nb_Votes), 20)) +
                 labs(x = "Number of votes") +
                 guides(fill = guide_legend(title = "Less than 20 votes"))
```


125
100
75
50
25
0
count
0
200
400
600
Number of votes
Less than 20 votes
FALSE
TRUE