Introductory analysis. Situation: people going inside and outside of a discoteque through two gates

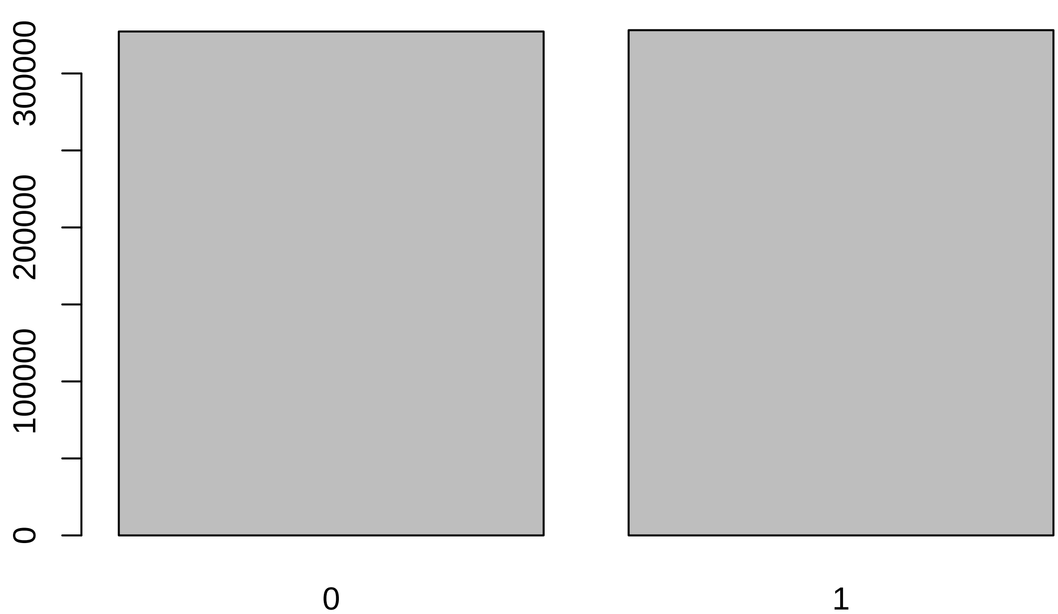
```r
data <- read.csv("data.csv")
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(tidyr)
barplot(table(data$receptor))
```



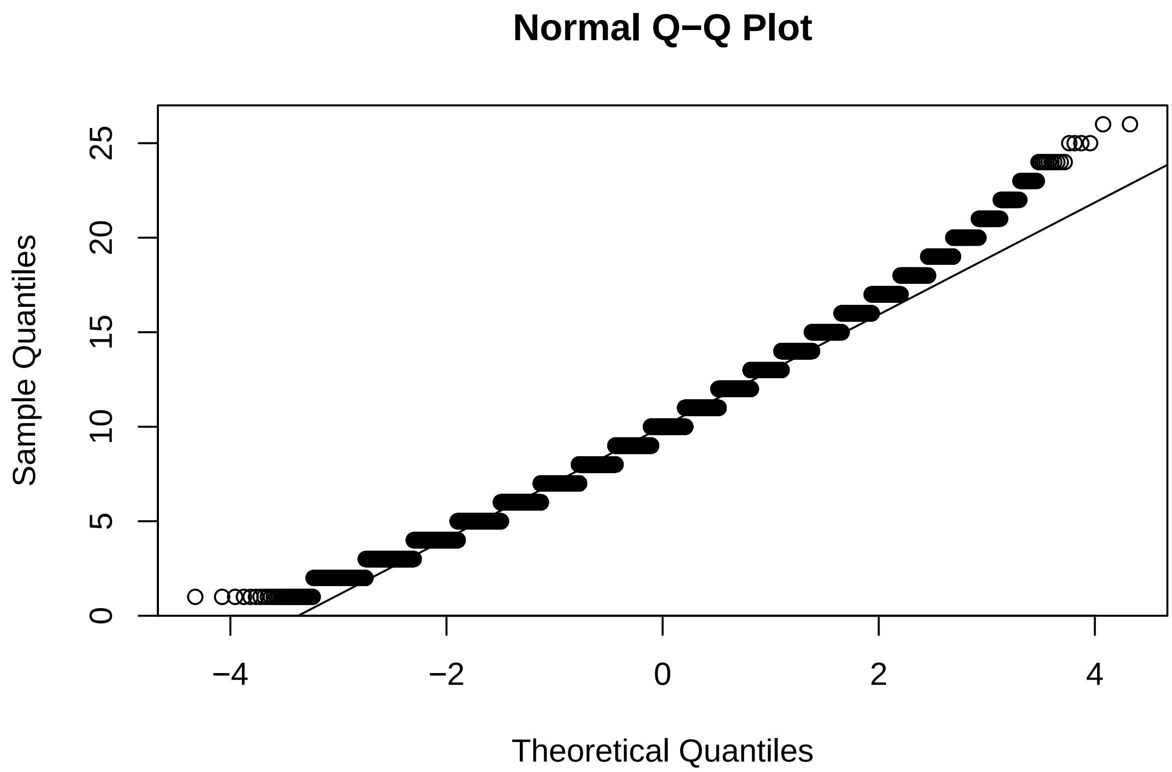

```r
num_macs <- count(data, MAC)$n
mean(num_macs)
```

```
## [1] 10.00092
```

```r
sd(num_macs)
```

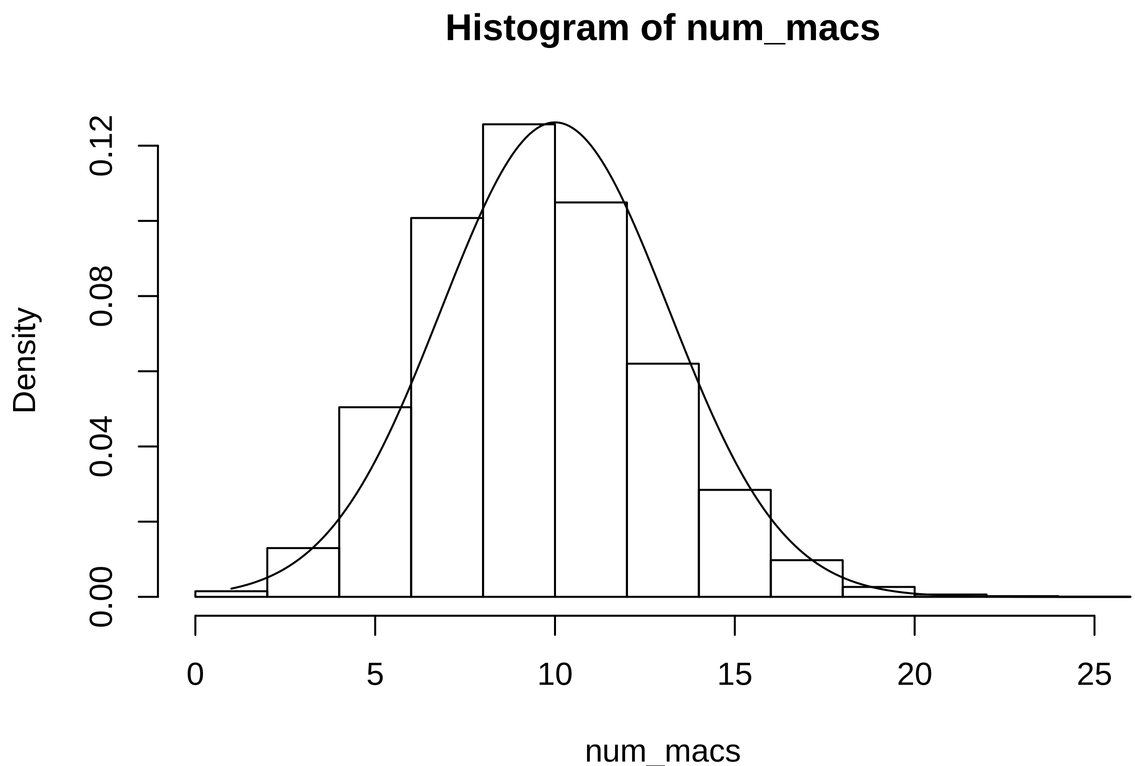
```
## [1] 3.16109
```

```r
qqnorm(num_macs)
qqline(num_macs)
```



It seems that the distribution the repetition of macs is a bit right-skewed. Seeing it in a histogramm, we can confirm this:

```
hist(num_macs, freq = F)
lines(seq(min(num_macs), max(num_macs), 0.1) ,dnorm(seq(min(num_macs), max(num_macs), by = 0.1),  mean =
```



Since this data is random and each timestamp a mac is generated with an uniform probability between all posibilities, and since the times a mac shows up is a sum of uniforms, due to the law of big numbers we can asume normality. TODO: normality test

Now: calculate who is inside and who is not. Objective:

MAC | t1 | t2 | t3

a | In | In | out

```
grouped <- data %>% group_by(MAC) %>%   mutate(
  is_in=as.logical(rank(timestamp) %% 2))
grouped[(grouped$MAC=="90b7"),]
```

```
## Source: local data frame [8 x 4]
## Groups: MAC [1]
##
##   timestamp    MAC receptor is_in
##       <int> <fctr>    <int> <lgl>
## 1         0   90b7        1  TRUE
## 2      8251   90b7        0 FALSE
## 3     98838   90b7        0  TRUE
## 4    158964   90b7        0 FALSE
## 5    179971   90b7        0  TRUE
## 6    202871   90b7        0 FALSE
## 7    353883   90b7        0  TRUE
```

```
## 8    619031   90b7          1 FALSE
```