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

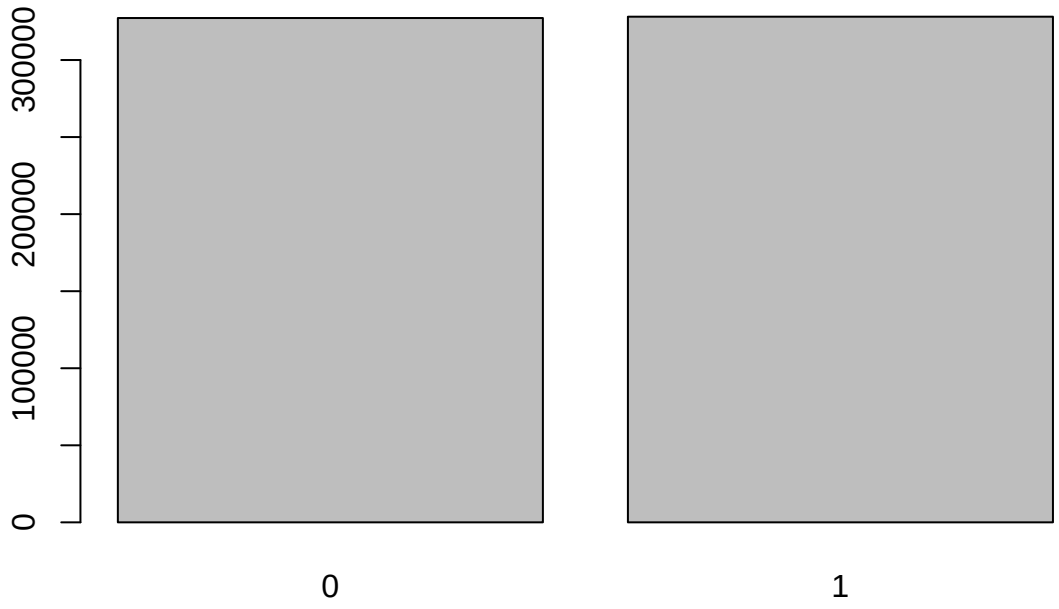
```
data <- read.csv("data.csv")
library(dplyr)
```

```
##
## 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
```

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



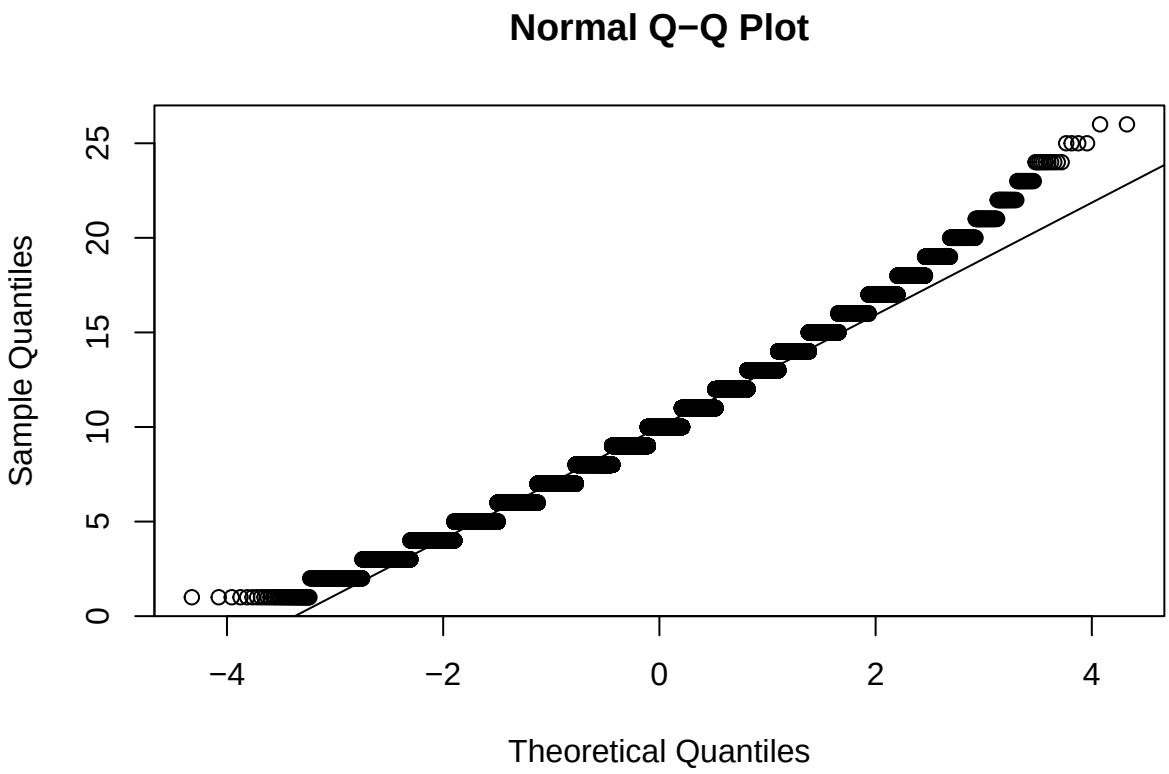

```
num_macs <- count(data, MAC)$n
mean(num_macs)
```

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

```
sd(num_macs)
```

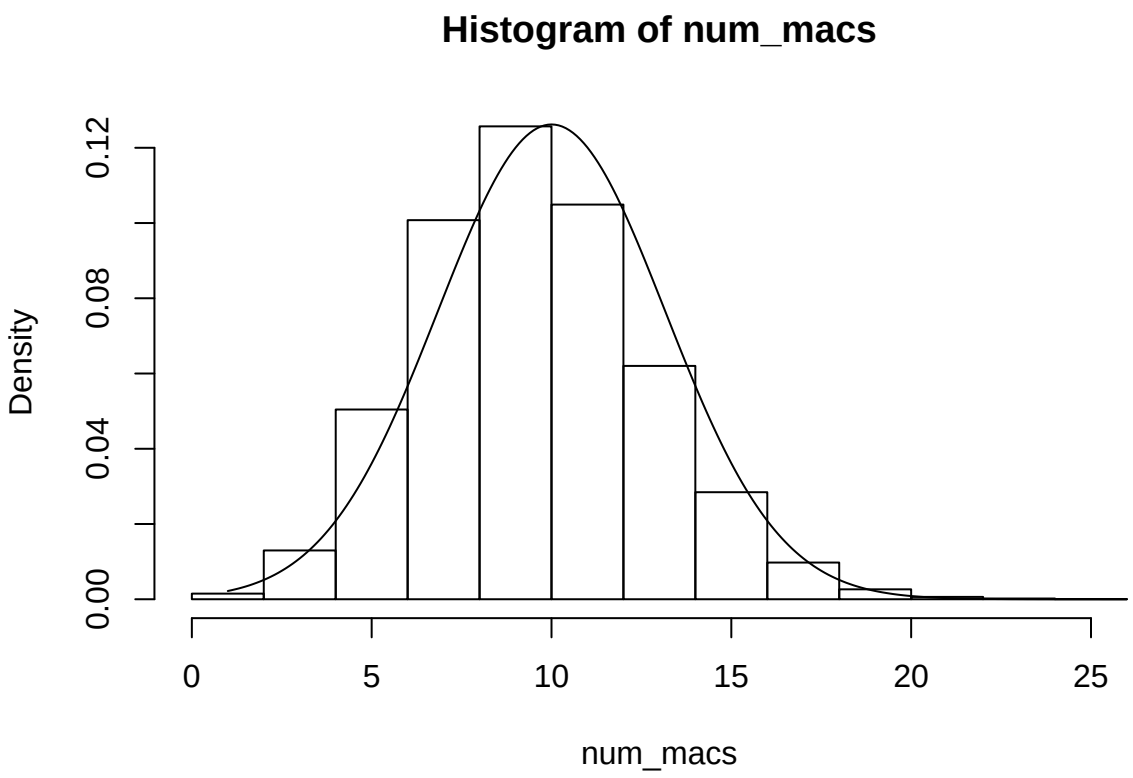
```
## [1] 3.16109
```

```
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:

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

```
## Source: local data frame [8 x 4]
## Groups: MAC [1]
##
##   timestamp    MAC receptor entering
##       <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
```

```
accumulated <- ungroup(grouped) %>% mutate(
  inside=cumsum(entering)-cumsum(!entering))
head(accumulated,40)
```

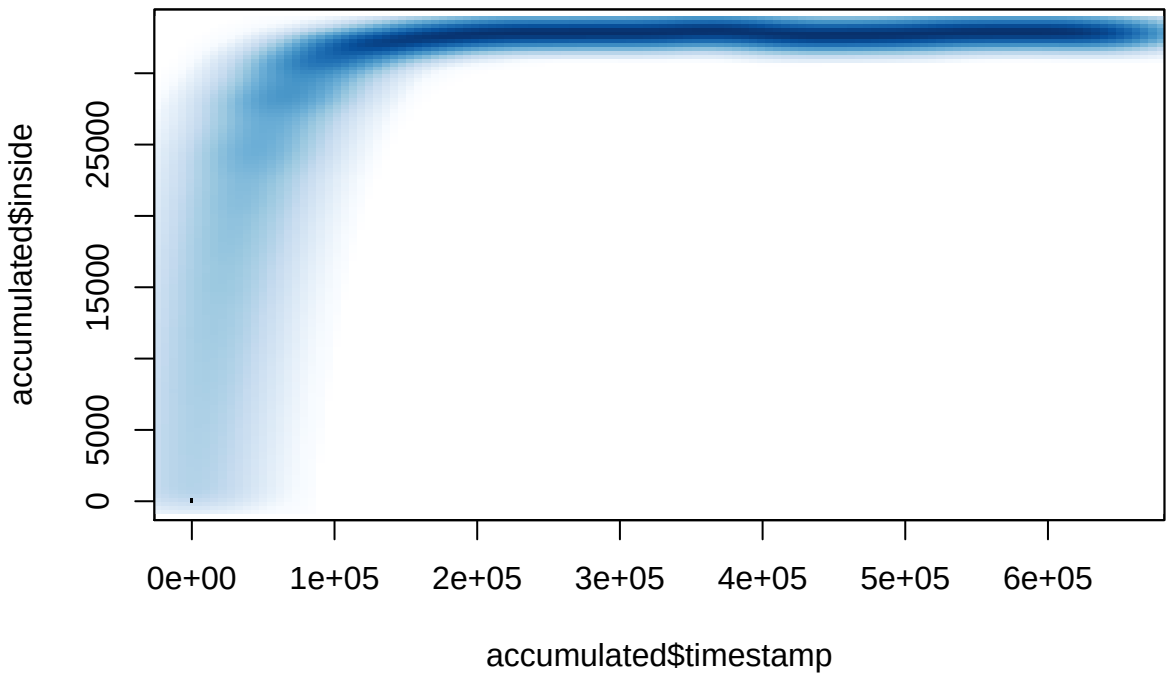
```
## # A tibble: 40 × 5
```

```
##    timestamp    MAC receptor entering inside
##        <int> <fctr>    <int>    <lgl>  <int>
## 1          0   90b7        1     TRUE      1
## 2          1   21b2        0     TRUE      2
## 3          2   25b3        1     TRUE      3
## 4          3   e9db        1     TRUE      4
## 5          4   aee2        0     TRUE      5
## 6          5   0538        0     TRUE      6
## 7          6   c11f        1     TRUE      7
## 8          7   4883        0     TRUE      8
## 9          8   5d0b        1     TRUE      9
## 10         9   cafe        0     TRUE     10
## # ... with 30 more rows
```

```
accumulated[(accumulated$MAC=="90b7"),]
```

```
## # A tibble: 8 × 5
##   timestamp    MAC receptor entering inside
##       <int> <fctr>    <int>    <lgl>  <int>
## 1         0   90b7        1     TRUE      1
## 2      8251   90b7        0    FALSE   7334
## 3     98838   90b7        0     TRUE  31199
## 4    158964   90b7        0    FALSE  32439
## 5    179971   90b7        0     TRUE  32702
## 6    202871   90b7        0    FALSE  32788
## 7    353883   90b7        0     TRUE  33094
## 8    619031   90b7        1    FALSE  32972
```

```
smoothScatter(accumulated$timestamp, accumulated$inside)
```


accumulated$inside
0
5000
15000
25000
0e+00
1e+05
2e+05
3e+05
4e+05
5e+05
6e+05
accumulated$timestamp