Part 1:

First, the birthdays can be formulated using numbers between 1 and 365 in a classroom with 10 people.

By applying probability approach, person 1 has a 365/365 or 1 while the person 2 has 364/365 and so forth. The chances of all people having an exclusive date of birth can be formulated:

$$1\times \frac{364}{365} \times \frac{363}{365} \times \frac{362}{365} \ . \ . \ . \ \frac{365-n+1}{365}$$

A function can be coded for any number in R:

```
n <-10

bd <- sample(1:365, n, replace = TRUE)
```

To determine which people of a vector have duplicates of elements, the function of 'duplicated' can be applied:

```
duplicated(c(1,2,3,4,2,3,5))
```

The function of 'any' and 'duplicated' are applied to check if two birthdays were the same:

```
any(duplicated(bd))
```

To estimate the probability with a condition of at least two people had the same birthday, the experiment is repeated over and over in given sampling sets:

```
N <- 10000

same_bd <-function(n){

  bd <-sample(1:365, n, replace = TRUE)

  any(duplicated(bd))

}
```

Part 2:

The outcomes are then obtained:

```
results <- replicate(N, same_bd(10))

mean(results, digits=3)
```

In addition, it can be proposed a function that calculates this in any group size:

```
comp_prob <- function(n, N=10000){

  results <- replicate(N, same_bd(n))

  mean(results)

}

mean(results, digits=3)
```

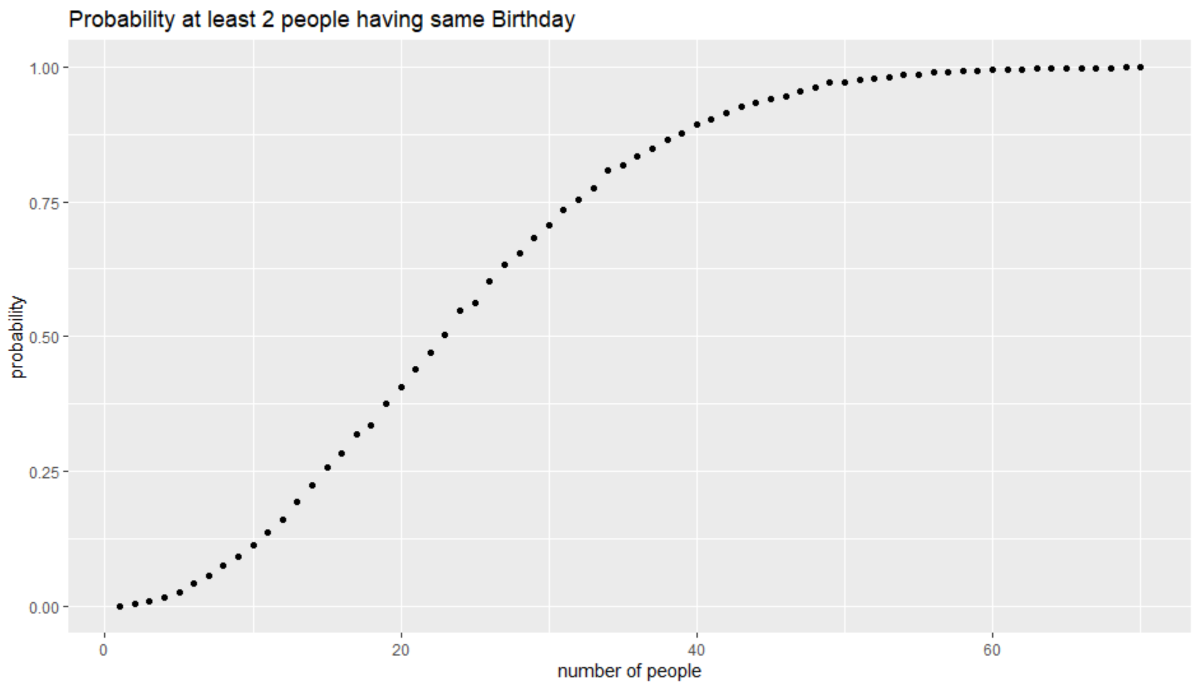
The probability exactly estimated is **0.117**.

Applying the function of 'sapply', the probability of element-wise operations can be obtained:

```
n <- seq(1,70)

prob <- sapply(n, comp_prob)
```

To illustrate it, a plot of the changes in a group of size can be gained:

```
library(tidyverse)

prob <- sapply(n, comp_prob)

qplot(n, prob, main="Probability at least 2 people having same Birthday",
xlab="number of people", ylab = "probability")
```



Part 3:

To make the experiments simpler, the multiplication rule can be applied to compute the probability of it that does not happen. The probability of 2 people having unique birthdays can be formulated:

$$1 - \frac{1}{365} = \frac{364}{365} = 0.997$$

23 people have 253 pairs.

$$\frac{23.22}{2} = 253$$

$$\left(\frac{364}{365}\right)^{253} = 0.499$$

The chance to find a match is:

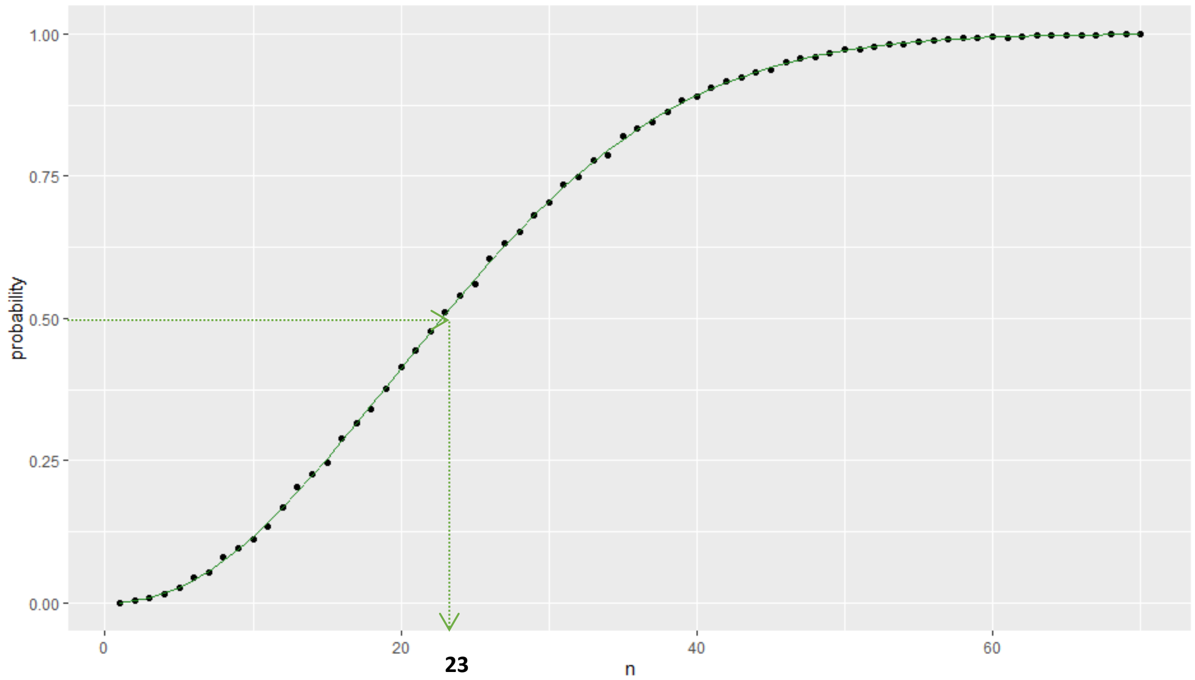
$$1 - 0.499 = 0.5001 = 50\%$$

Finally, the formula can be obtained:

$$p(n) = 1 - \left(\frac{364}{365}\right)^{C(n,2)} = 1 - \left(\frac{364}{365}\right)^{n(n-1)/2}$$

According to that equation, a function can be coded in R:

```
est_prob <-function(n){

  prob_bd <-seq(365, 365-n+1)/365

  1-prod(prob_bd)

}

eprob <- sapply(n, est_prob)

qplot(n, prob, xlab="number of people", ylab = "probability") +
geom_line(aes(n, eprob), col="forestgreen")
```



Using 10,000 experiments, a function can be coded as follow:

```
N <- 10^seq(1, 4, len = 100)

est_prob <- function(N, n=23){

  same_day <- replicate(N, same_bd(n))

  mean(same_day)

}

prob <- sapply(N, est_prob)

qplot(log10(N), prob, main="The chance of 50% for n=23 people",
geom="line")+geom_point()
```

For a group of **23 people** with 10,000 experiments, a plot of the **probability showing 50**% can be described as follow: