Project 3

Natalie Sullivan Wu

In this project, I plan to simulate the probability of a shared birthday. In a room of N amount of people, I will see what the probability of at least two people sharing the same birthday is. There is a theory called the birthday paradox, where the probability of two people sharing the same birthday increases with the number of possible pairings, not just the group size.

```
library(tidyverse)
library(purrr)
library(ggplot2)
```

First we will find the birthday paradox for N amount of people.

```
sim_shared_bday <- function(N){
  bdays <- sample(1:365, N, replace = TRUE)
  return(length(bdays) != length(unique(bdays)))
}
```

Next we will estimate the probability of shared birthdays. We will use multiple simulations using sapply() to make sure our estimation is accurate. sapply() will apply the function sim_shared_bday to the simulation multiple times

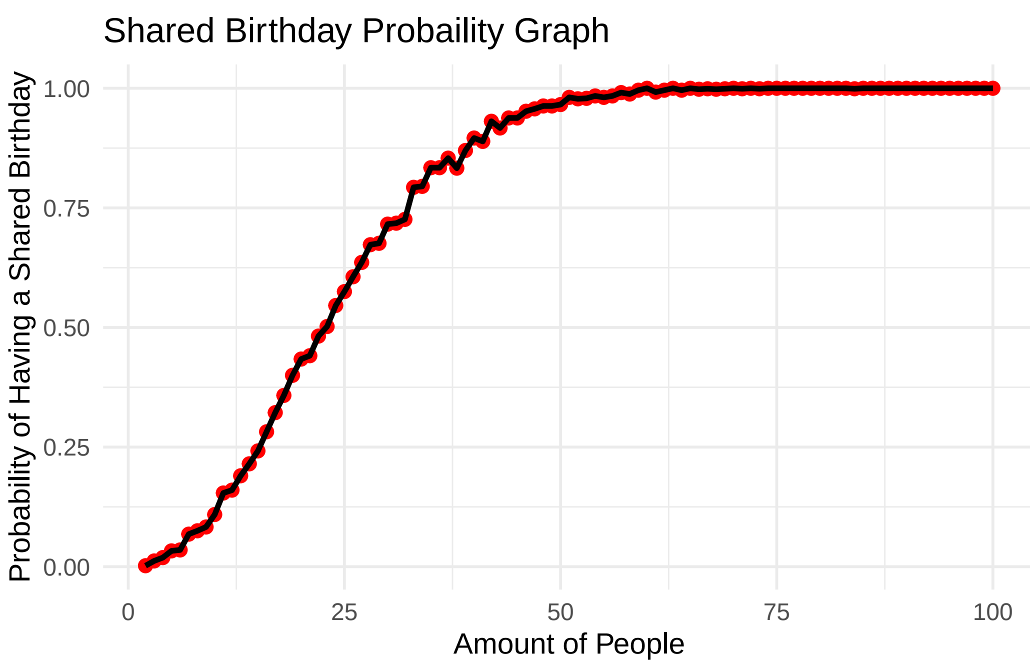
```
estimate_shared_bday <- function(N, num_sims){
  outcome <- sapply(1:num_sims, function(x) sim_shared_bday(N))
  return(mean(outcome))
}
```

Now that we have an estimate of our probability, we will set out parameters and define our probability. I set the group size from 2 to 100 people, and set the number of simulations to 1000 so get more accurate and precise results. I am using map() to apply the estimate_shared_bday to different group sizes. Lastly, I will plot my results.

In the plot, the point represent the probability that relate to each group size. The line represents the probability of at least two people sharing the same birthday, as the group size increased.

```r
grp_size <- 2:100
num_sims <- 1000
prob <- map_dbl(grp_size, ~ estimate_shared_bday(.x, num_sims))
data <- data.frame(group_size = grp_size, probability = prob)

ggplot(data, aes(x = group_size, y = probability)) +
  geom_point(color = "red", size = 2) +
  geom_line(size = 1) +
  labs(
    title = "Shared Birthday Probaility Graph",
    x = "Amount of People",
    y = "Probability of Having a Shared Birthday"
  ) +
  theme_minimal()
```

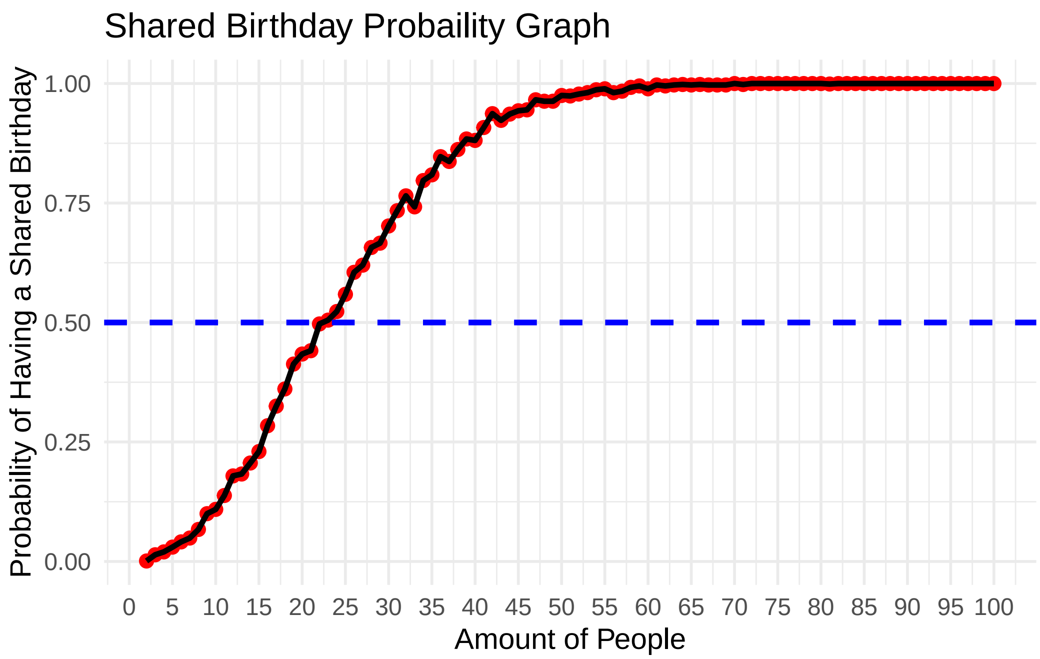


From this plot, we can see that as the group size increases, there is a rapid increase in probability. It is an exponential increase in probability as we increase group size. This contracts what we may have intuitively thought. We naturally would have thought that the probability would increase slowly, however the graph clearly shows it does not. Additionally, the graph shows how after the exponential growth, the slope drastically decreases around a group size of 50 people.

I want to see where the probability of sharing a birthday with another person is 50%, so I will add a horizontal line to show the intersection.

```r
grp_size <- 2:100
num_sims <- 1000
prob <- map_dbl(grp_size, ~ estimate_shared_bday(.x, num_sims))
data <- data.frame(group_size = grp_size, probability = prob)

ggplot(data, aes(x = group_size, y = probability)) +
  geom_point(color = "red", size = 2) +
  geom_line(size = 1) +
  geom_hline(yintercept = 0.5, linetype = "dashed", color = "blue", size = 1) +
  scale_x_continuous(breaks = seq(0, 100, by = 5)) +
  labs(
    title = "Shared Birthday Probaility Graph",
    x = "Amount of People",
    y = "Probability of Having a Shared Birthday"
  ) +
  theme_minimal()
```



We can see that the point of interception is around the group size of 23. This means that in a group of 23 people, there would be a 50% probability of at least two people sharing a birthday within the group.

For this project, I basically found the probability that at least two people share the same birthday within a group of people. The size of the group changes, and with that there is an exponential increase in the probability. To achieve this, I created two functions that simulate and estimate the probability then graphed it with the y-axis being the probability and the x-axis being the amount of people with a min of 0 and a max of 100, however, the function starts with a minimum of 2 people. From this, we were able to see how probability interacts with number of people when finding the probability of at least two people sharing birthdays.