## Question 1

- This is a well known problem known as The Coupon collector's Problem.
  - Estimating the expected number of dice rolls using R

```
## function that simulates rolling a n-sided die
diceroll <- function(n) {
  die <- 1:n
  result <- sample(die, 1)
  result
}
## function that simulates rolling a 12-sided die until all possible outcomes
## have happened at least once
rollcollector <- function() {
  rolls <- diceroll(12)
  result <- 1
  while(length(unique(rolls))<12) {
    rolls <- append(rolls, diceroll(12))
    result <- result + 1
  }
  result
}
```

    * Expected value of 100 trials:

```
> trials <- 100
> simlist <- replicate(trials, rollcollector())
> mean(simlist)
[1] 40.22
```

    * Expected value of 1000000 trials:

```
> trials <- 1000000
> simlist <- replicate(trials, rollcollector())
> mean(simlist)
[1] 37.24014
```

  - In order to solve this problem analytically, we will define the following random variables: $X_i$: number of dice rolls needed until the $i^{th}$ outcome has occurred

$$X_i = X_1 + X_2 + X_3 + \cdots$$

$Y_i$: number of dice rolls between the $(i-1)^{th}$ occurring for the first time and the $i^{th}$ outcome occurring for the first time

$$Y_i = X_{i+1} - X_i$$

Our goal is to compute $E(X_{12})$, the expected number of dice rolls until all possible outcomes of our 12-sided die have occurred at least once. Using the definitions above, we can see that

$$X_{12} = \sum_{i=0}^{11} Y_i$$

This will prove to be useful soon enough.
Now, let's find the P.M.F. of $Y_i$, starting with finding the probability of $Y_1$. There is a total of 12 possible outcomes, and we have already seen 1. So there are $12 - 1$ potential sides left. Therefore

$$Y_1 = Geo(\frac{12-1}{12})$$

Generalising this, we get

$$Y_i = Geo(\frac{12-i}{12})$$

Since $Y_i$ follows a geometric distribution, we know that $E(Y_i)$ is simply the inverse of the P.M.F.

$$E(Y_i) = \frac{12}{12-i}$$

Now we can compute $E(X_{12})$

$$E(X_{12}) = E[\sum_{i=0}^{11} Y_i]$$

$$= \sum_{i=0}^{11} E(Y_i)$$

$$= \sum_{i=0}^{11} \frac{12}{12-i}$$

$$= 12 \sum_{i=0}^{11} \frac{1}{12-i}$$

$$= \frac{86021}{2310}$$

$$\approx 37.2385$$

- Having already completed the first part of the problem, this part is quite simple as we only need to expand our code to include two dice

```
## expanding rollcollector
rollcollector2 <- function() {
  rolls <- sum(replicate(2, diceroll(6)))
  result <- 1
  while(length(unique(rolls))<11) {
    rolls <- append(rolls, sum(replicate(2, diceroll(6))))
    result <- result + 1
  }
  result
}
trials <- 10000
simlist2 <- replicate(trials, rollcollector2())
```

- Expected value of 100 trials

```
> trials <- 100
> simlist2 <- replicate(trials, rollcollector2())
> mean(simlist2)
[1] 62.77
```

- Expected value of 100000 trials

```
> trials <- 100000
> simlist2 <- replicate(trials, rollcollector2())
> mean(simlist2)
[1] 61.27343
```