Module 3 Peer Review Assignment

In [ ]:

```

```

Problem 1

You work at a factory that manufactures light bulbs. You have determined that 5% of light bulbs that are produced are defective. For each of the scenarios below:

1. Define an appropriate random variable and distribution.
2. State the values that the random variable can take on.
3. State any assumptions that you need to make.
4. Find the probability that the random variable you defined takes on the value $X = 4$.

Part a)

Out of 30 lightbulbs, k are defective.

1. $X \sim B(30, 0.05)$, Binomial
2. X can take on following values $\{0, 1, 2, \ldots, 29, 30\}$
3. X can be represented as sum of independently identically distributed Bernoulli Random Variables (independent trials for the Binomial), $0 \leq k \leq 30$
4. $P(X = k) = \binom{30}{k}(0.05)^k(0.95)^{(30-k)}$, hence, $P(X = 4) = \binom{30}{4}(0.05)^4(0.95)^{26} = 0.04513605$

Part b)

You test each lightbulb as it comes of the line. The k^{th} light bulb is the first defective light bulb you find.

1. $X \sim Geom(0.05)$, Geometric
2. X can take on following values $\{0, 1, 2, \ldots, \}$
3. $k \geq 0$
4. $P(X = k) = (1 - 0.05)^{k-1} \times 0.05$, hence,
   $P(X = 4) = 0.95^3 * 0.05 = 0.04513605 = 0.04286875$

Part c)

You find your second defective light bulb after observing k light bulbs in all.

1. $X \sim NB(2, 0.05)$, Negative Binomial with 2 successes (defects)
2. X can take on following values $\{2, \ldots, \}$, number of failures before 2^{nd} success (defect)
3. We have independent Bernoulli trials, with $k \geq 2$, we need to find probability of $k - 2$ failures before the 2^{nd} success
4. $P(X = k) = \binom{2+k-2-1}{2-1}(0.05)^2(0.95)^{k-2} = \binom{k-1}{2-1}(0.05)^2(0.95)^{k-2}$. Hence, probability to find the second defective light bulb after observing $k = 4$ light bulbs in all =
   $P(X = 4) = \binom{3}{1}(0.05)^2(0.95)^2 = 0.00676875$

Problem 2

Consider a loaded six-sided die that is twice as likely to roll an even number as an odd number. Let X be random variable for value that is rolled from the die.

Part a)

What is the Probability Mass Function for X. Write this out as a table.

Let $P(\text{X is odd}) = p$. Given $P(\text{X is even}) = 2.P(\text{X is odd}) = 2p$, we have $3.2p + 3.p = 1 \implies p = \frac{1}{9}$. Hence, the PMF is given below:

x	$P(X = x)$
1	$1/9$
2	$2/9$
3	$1/9$
4	$2/9$
5	$1/9$
6	$2/9$

Part b)

What is the Cumulative Distribution Function for X?

CDF of X is $F(X \leq a) = \sum\limits_{x \leq a} P(X = x)$, the CDF is shown in the given table:

X	$F(X)$
$1 \geq X$	0
$1 < X \leq 2$	$1/9$
$2 < X \leq 3$	$3/9$
$3 < X \leq 4$	$5/9$
$4 < X \leq 5$	$6/9$
$5 < X \leq 6$	$7/9$
$X \geq 6$	1

Part c)

What is $E[X]$?

$$E[X] = \sum_{x=1}^{6} x.\, P(X=x) = \frac{1}{9}(1+3+5) + \frac{2}{9}(2+4+6) = 3.6667$$

In [1]:

```
x <- 1:6
p <- c(1/9,2/9,1/9,2/9,1/9,2/9)
sum(p*x) # E[X]
```

3.66666666666667

Problem 3

How would we simulate variables from these distributions in R? It'll turn out that the method is fairly similar across all these distributions so, for simplicity, let's just say we want to simulate $X \sim Bin(n,p)$. Take a look at the official documentation for this function here (https://www.rdocumentation.org/packages/stats/versions/3.3/topics/Binomial). Not extremely clear, is it? Let's go through it one step at a time.

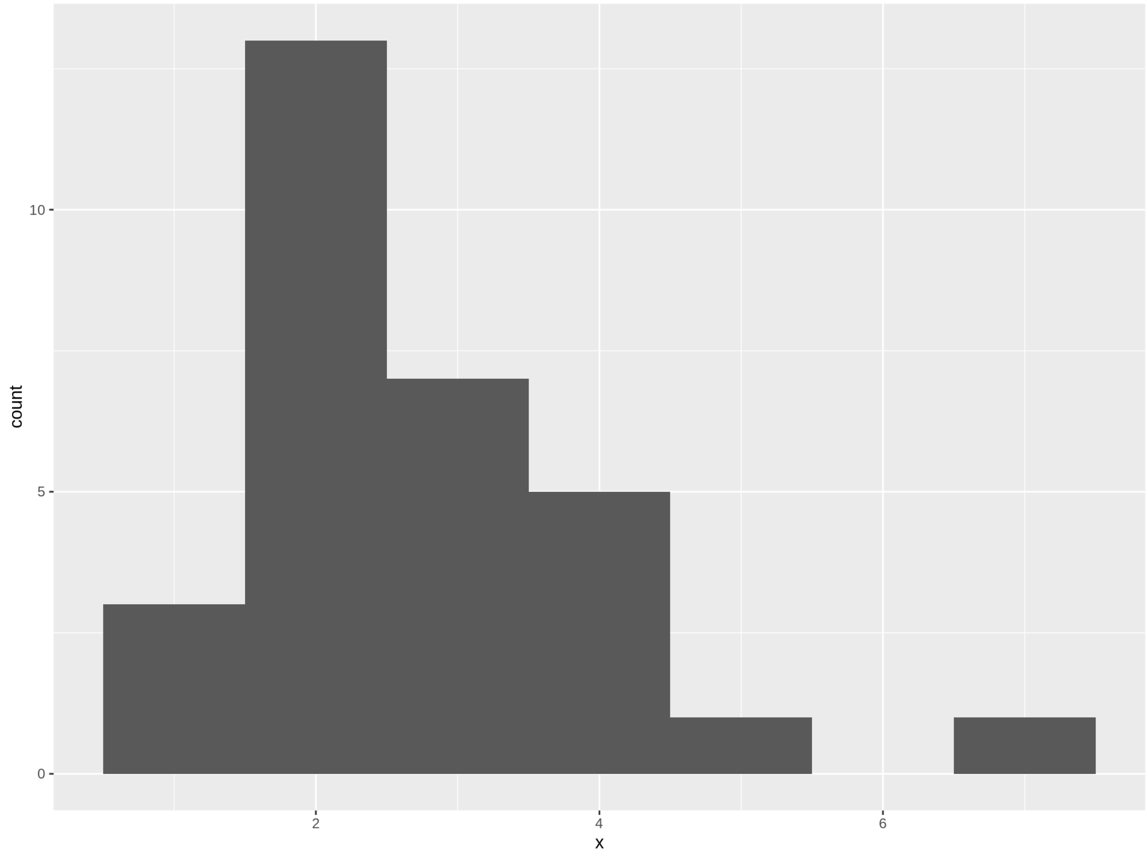
Part a)

What if we want a random variable from this distribution? That is, we know some underlying distribution and we want to simulate many results from that distribution. Then we would use the "random generation" function `rbinom()` .

Play around with this function, with different `size` and `prob` parameters to get a feel for how it works. Finally, generate 30 results from a $Bin(10, 0.3)$ distribution and plot a histogram of the results.

In [10]:

```
x <- rbinom(n=30, size=10, prob=0.3)
library(ggplot2)
options(repr.plot.width = 10, repr.plot.height = 7.5, repr.plot.res = 200)
ggplot() + geom_histogram(aes(x), binwidth=1)
```



Part b)

What if we have some value k and we want to know what's probability of generating k? That is, we're solving the Probability Mass Function $P(X = k)$. Then we would use the "density" function `dbinom()` .

Let $X \sim Bin(15, 0.4)$. By hand, solve $P(X = 4)$. Then use the `dbinom()` function to confirm your result.

$$P(X = 4) = \binom{15}{4}(0.4)^4(0.6)^{11} = 0.1267758$$

In [12]:

```
dbinom(4, size=15, prob=0.4)
```

0.12677580324864

Part c)

What if we wanted to solve for some value of the Cumulative Density Function? That is, we know k and want to find $P(X \leq k) = p$. Then we would use the "distribution function" `pbinom()` .

Let $X \sim Bin(15, 0.4)$. By hand, solve $P(X \leq 4)$. Then use the `pbinom()` function to confirm your result.

$$P(X \leq 4) = \sum_{k=0}^{4} \binom{15}{k}(0.4)^k(0.6)^{15-k} = 0.2172777$$

In [14]:

```
pbinom(4, 15, 0.4)
```

0.217277705650176

Part d)

Finally, we have the "quantile" function `qbinom()` . This function is the reverse of the `pbinom()` function, in that it takes a probability p as an argument and returns the value k of the CDF that results in that much probability.

Use the `qbinom()` function to confirm your results from **Part c**. That is, plug in the probablity you got from **Part c** and see if you get the same k.

In [16]:

```
qbinom(0.2172777, 15, 0.4)
```

4

Nearly every distribution has these four functions, and they will be very useful for our future calculations and simulations.