Activity 4

Name here

Q1.

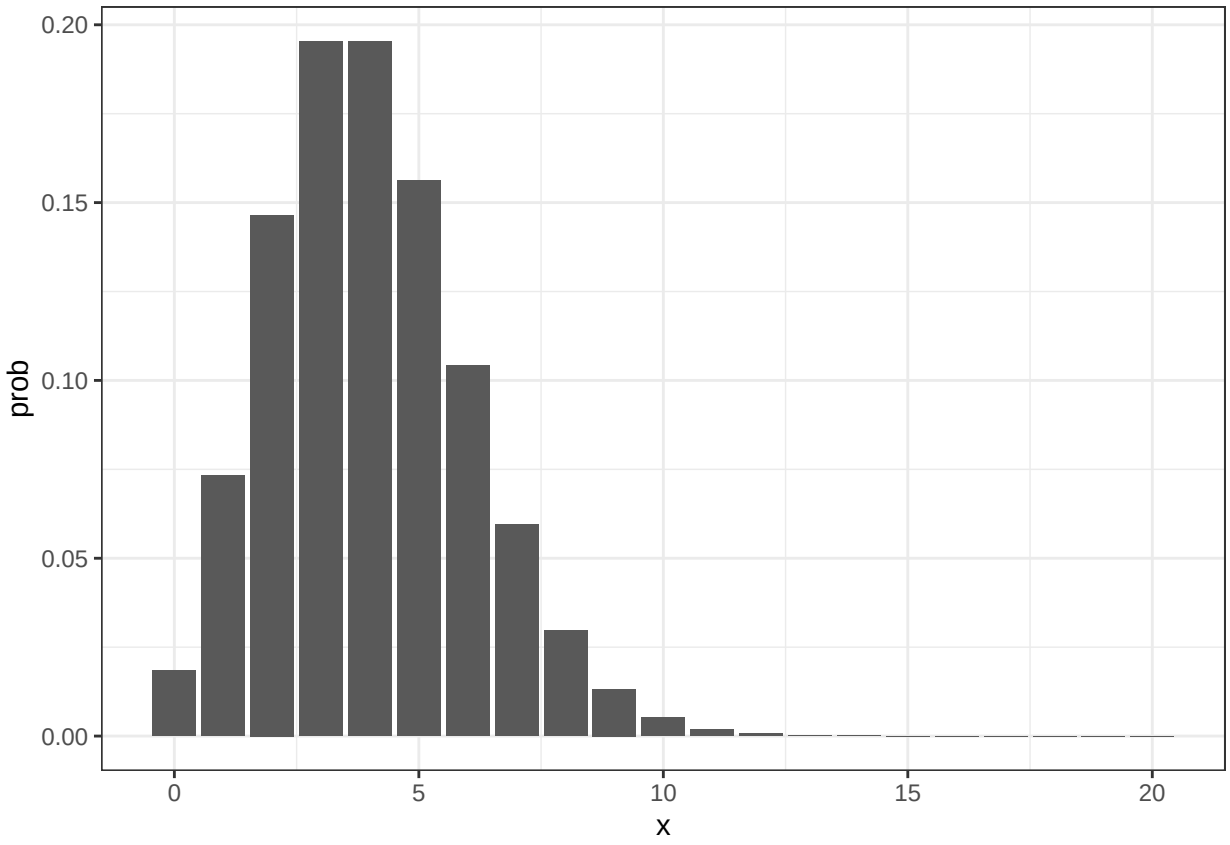
Describe the differences between probability density functions and probability mass functions.

Q2.

The highest density interval (HDI) corresponds to the shortest possible interval, for a specified probability level. What would be the 95% HDI for the following discrete distributions?

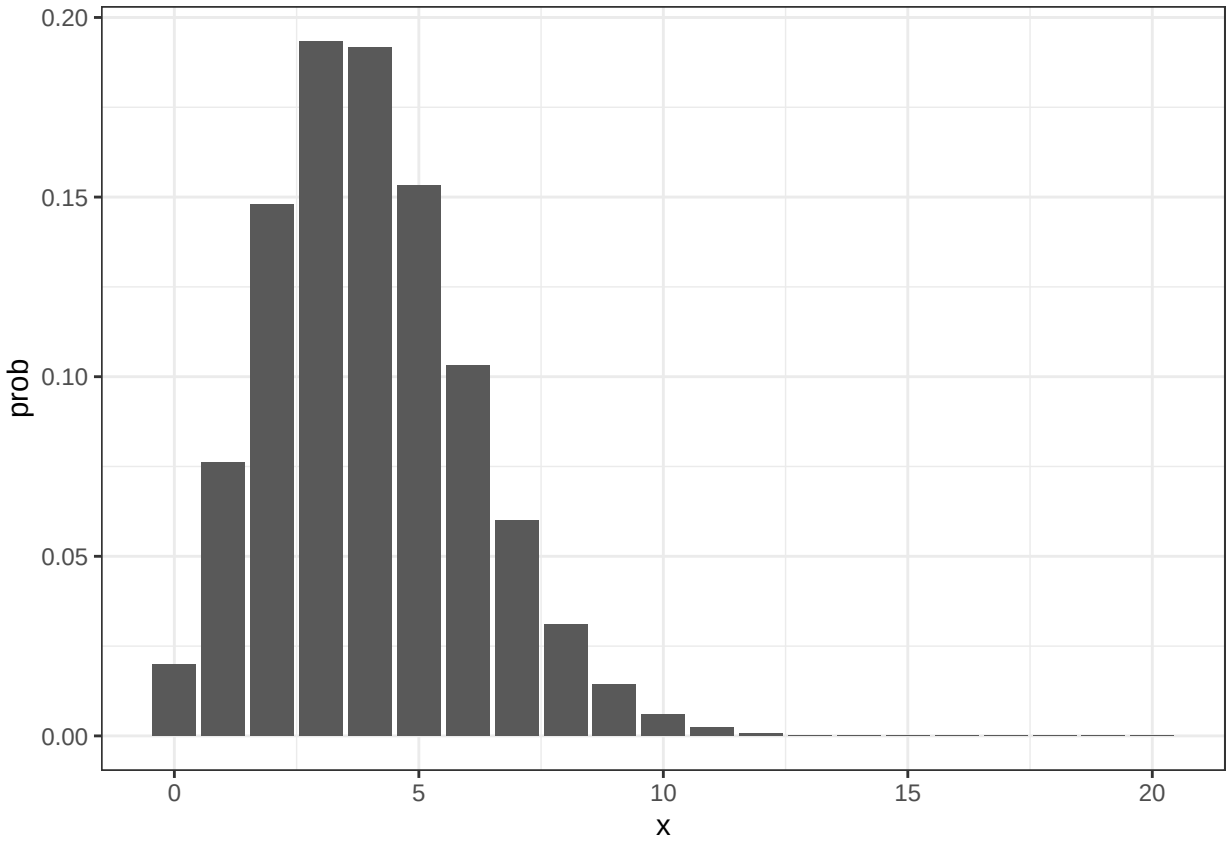
a. Poisson(4)

```
library(tidyverse)
x_seq <- 0:20
tibble(x = x_seq, prob = dpois(x_seq,4)) %>%
  ggplot(aes(x=x,y=prob)) + geom_col() + theme_bw()
```



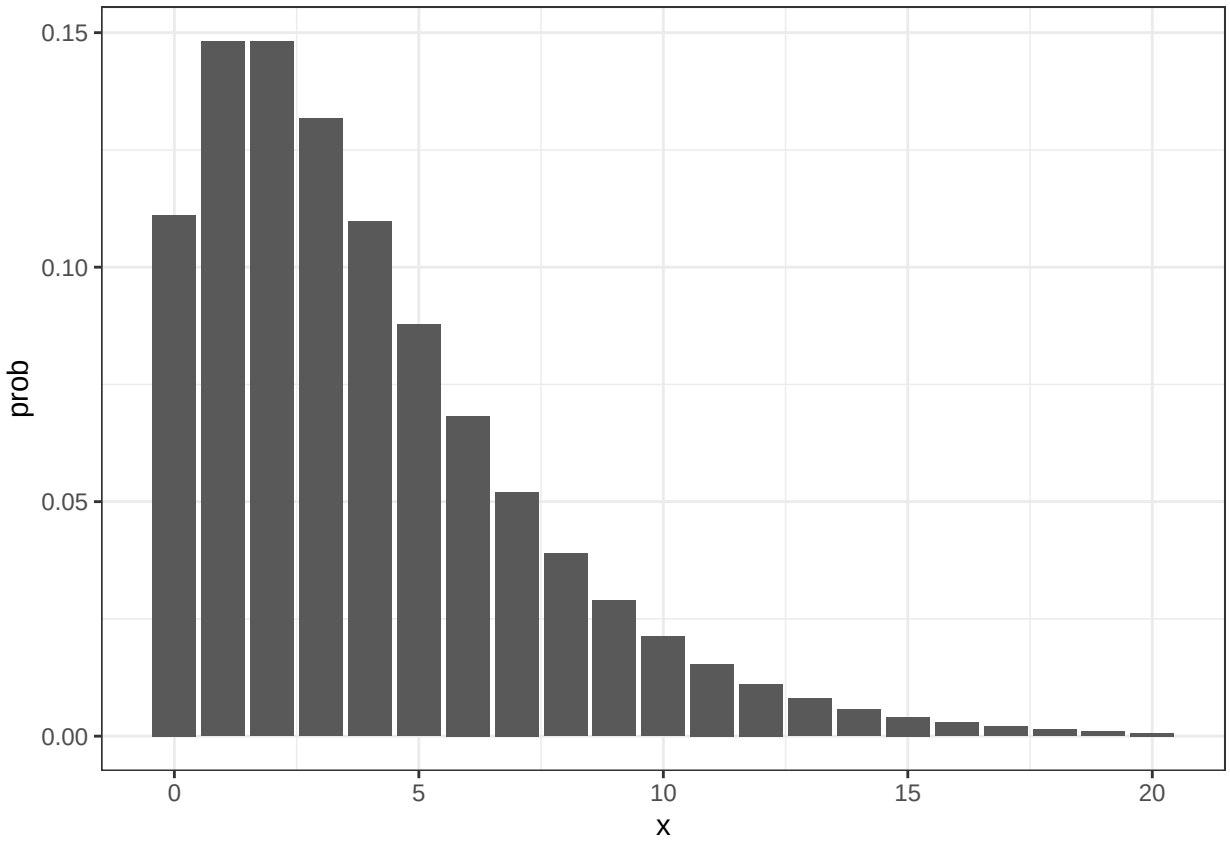
b. NB(4, 100)

```
x_seq <- 0:20
tibble(x = x_seq, prob = dnbinom(x_seq,mu =4, size = 100)) %>%
  ggplot(aes(x=x,y=prob)) + geom_col() + theme_bw()
```



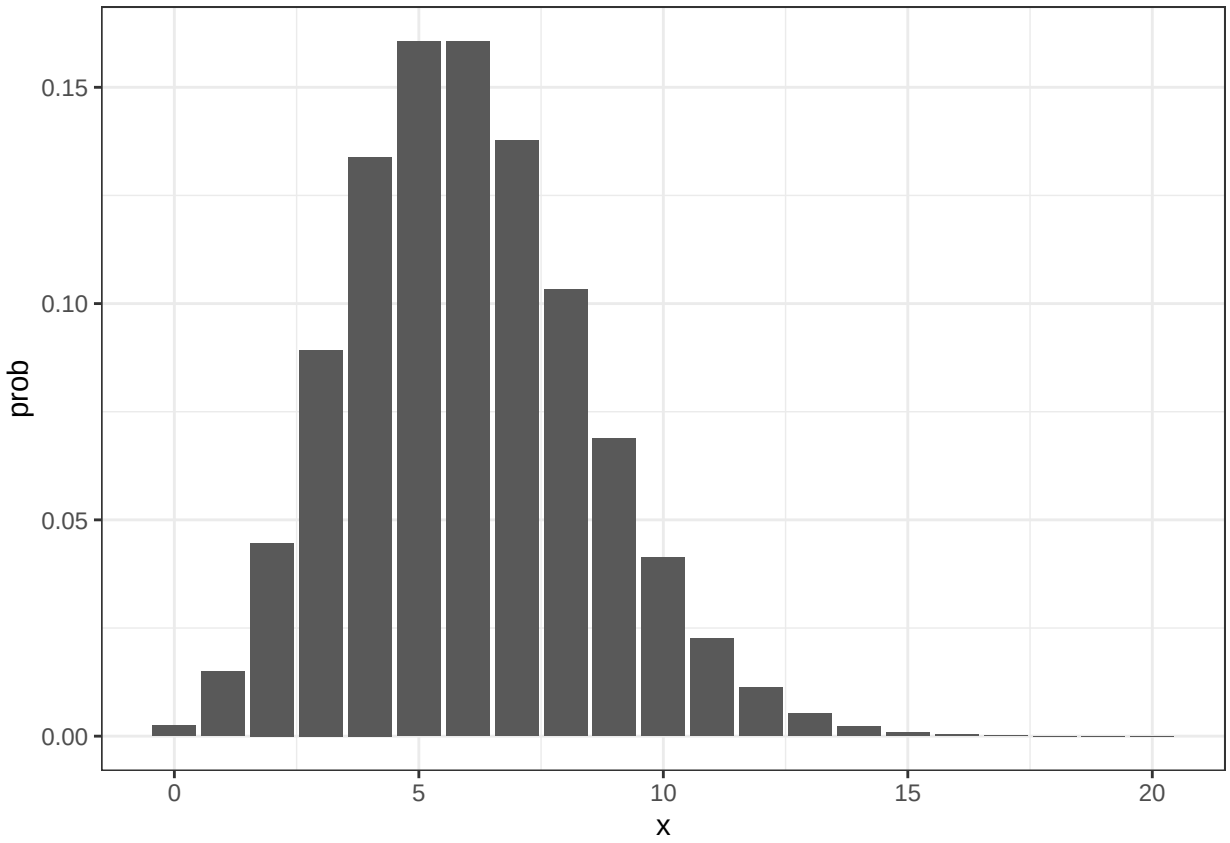
d. NB(4, 1)

```
x_seq <- 0:20
tibble(x = x_seq, prob = dnbinom(x_seq,mu =4, size = 2)) %>%
  ggplot(aes(x=x,y=prob)) + geom_col() + theme_bw()
```



2. Pois(6)

```
tibble(x = x_seq, prob = dpois(x_seq,6)) %>%
  ggplot(aes(x=x,y=prob)) + geom_col() + theme_bw()
```



This would be approximately (2, 11) (`sum(dpois(2:11,6))`)

Q3.

Use a dataset containing homes in the Seattle, WA area http://www.math.montana.edu/ahoegh/teaching/stat408/datasets/SeattleHousing.csv for this question.

```
seattle <- read_csv('http://www.math.montana.edu/ahoegh/teaching/stat408/datasets/SeattleHousing.csv') %
  mutate(bath_category = case_when(
    bathrooms <= 2 ~ '0 - 2',
    bathrooms > 2 & bathrooms <=4 ~ '2 - 4',
    bathrooms > 4 ~ 'more than 4'),
    bed_category = case_when(
    bedrooms == 0 ~ '0',
    bedrooms > 0 & bedrooms <= 2 ~ '1-2',
    bedrooms > 2 & bedrooms <= 4 ~ '3-4',
    bedrooms > 4 ~ 'more than 5'
   )
  )
```

```
## Rows: 869 Columns: 14
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (14): price, bedrooms, bathrooms, sqft_living, sqft_lot, floors, waterfr...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Use the two-by-two table containing bathrooms (grouped as: 0 - 2, more than 2 - 4, more than 4) and bedrooms (0, 1-2, 3-4, 5 or more).

```
table(seattle$bath_category, seattle$bed_category)
```

```
##
##                 0 1-2 3-4 more than 5
##   0 - 2         3 129 332          13
##   2 - 4         0  22 314          32
##   more than 4   0   0  10          14
```

a. Compute marginal probability of having 0 - 2 bathrooms

b. Compute joint probability of having 0 - 2 bathroom and 3-4 bedrooms

c. Compute conditional probability of having 3-4 bedrooms given that it has 0 - 2 bathrooms

d. Are bathrooms and bedrooms independent? Why or why not.

Q4. (DBDA 4.5B)

Use a normal curve to describe the following belief. Suppose you believe that women's heights follow a bell-shaped distribution, centered at 162 cm with about 2/3rds of all women having heights between 147 and 177. What should be the μ and σ values?

Q5.

Assume your roommate has taken, and tested positive, for an at home antigen Covid test. Let's assume the test was from Cochrane (data)[https://www.cochrane.org/CD013705/INFECTN_how-accurate-are-rapid-antigen-tests-diagnosing-covid-19] with the following properties:

- P[Test = + | Disease = +] = .82
- PP[Test = - | Disease = -] = .995

a. If the overall population prevalence was 5%, what is the probability that your roommate has COVID-19?

b. Now assume that your other roommate took a test (with the same characteristics) and that test was negative. What is the probability the other roommate has COVID-19?