This document will describe how to use R to calculate probabilities associated with common distributions as well as to graph probability distributions. R has a number of built in functions for calculations involving probability distributions, both discrete and continuous. We have already seen the binomial, Poisson and the normal distributions during our introduction to probability. Later, we will use the t, F, and chi-square distributions in our study of statistical inference.

For each of these distributions (and others), R has four primary functions. Each function has a one letter prefix followed by the root name of the function. The names make mnemonic sense for continuous random variables but are used in both cases. For example `dnorm` is the height of the **d**ensity of a normal curve while `dbinom` returns the probability of an outcome of a binomial distribution. Here is a table of these commands.

	Meaning	
Prefix	Continuous	Discrete
d	**d**ensity	probability (pmf)
p	**p**robability (cdf)	probability (cdf)
q	**q**uantile	**q**uantile
r	**r**andom	**r**andom

Distribution	Root
Binomial	`binom`
Poisson	`pois`
Normal	`norm`
t	`t`
F	`F`
Chi-square	`chisq`

A useful tip. The arrows of the keyboard (up and down) recall previous commands. It's easy to modify previous long commands without typing them in all again.

The Binomial Distribution. The binomial distribution is applicable for counting the number of outcomes of a given type from a prespecified number n independent trials, each with two possible outcomes, and the same probability of the outcome of interest, p. The distribution is completely determined by n and p. The probability mass function is defined as:

$$\Pr\{Y = j\} = \binom{n}{j} p^j (1-p)^{n-j}$$

where

$$\binom{n}{j} = \frac{n!}{j!(n-j)!}$$

is called a binomial coefficient. (Our textbook uses the notation ${}_nC_j$ instead.) In R, the function `dbinom` returns this probability. There are three required arguments: the value(s) for which to compute the probability (j), the number of trials (n), and the success probability for each trial (p).

For example, here we find the probability of exactly 2 successes in 5 trials with $p = 0.1$

```
> dbinom(2, 5, 0.1)
[1] 0.07290
```

If we want to find the complete distribution when $n = 5$ and $p = 0.1$, we do this.

```
> dbinom(0:5, 5, 0.1)
[1] 0.59049 0.32805 0.07290 0.00810 0.00045 0.00001
```

The function `pbinom` is useful for summing consecutive binomial probabilities. With $n = 5$ and $p = 0.1$, here are some example calculations.

```
> sum(dbinom(0:2, 5, 0.1))
[1] 0.99144
> pbinom(2, 5, 0.1)
[1] 0.99144
```

It means that

$$\begin{aligned} \Pr\{Y \leq 2\} &= \text{pbinom}(2,5,0.1) = 0.99144 \\ \Pr\{Y \geq 3\} &= 1 - \Pr\{Y \leq 2\} = 1 - \text{pbinom}(2,5,0.1) = 0.00856 \\ \Pr\{1 \leq Y \leq 3\} &= \Pr\{Y \leq 3\} - \Pr\{Y \leq 0\} = \text{pbinom}(3,5,0.1) - \text{pbinom}(0,5,0.1) = 0.40905 \end{aligned}$$

We can also find the quantiles of a binomial distribution. For example, the following command finds the 90th percentile of a binomial distribution with $n = 200$ and $p = 0.3$.

```
> qbinom(0.9, 200, 0.3)
[1] 68
```

It means that the probablity that the outcome if less than 68 (less than or equal to 68?) is approximately 90%. Let us check that out.

```
> pbinom(67, 200, 0.3); pbinom(68, 200, 0.3)
[1] 0.8757949
[1] 0.9040488
```

The last function for the binomial distribution is used to take random samples. Here is a random sample of 20 binomial random variables drawn from the binomial distribution with $n = 10$ and $p = 0.5$.

```
> rbinom(20, 10, 0.5)
[1] 6 7 3 5 3 6 7 6 5 8 5 5 6 4 5 7 5 3 5 6
```

Graphing Probability Distributions. Graphs is one the best way to communicate information. R makes beautiful graphs, highly customizable, but on the other hand there are not always easy to do. I will first give you a glance at the basics commands to plot graphs. A plot is built with the command `plot`. Many other things (lines, points, colors, titles and sub-titles, axes, legends, etc.) can then be added to the plot with other commands. I suggest that you try the following.

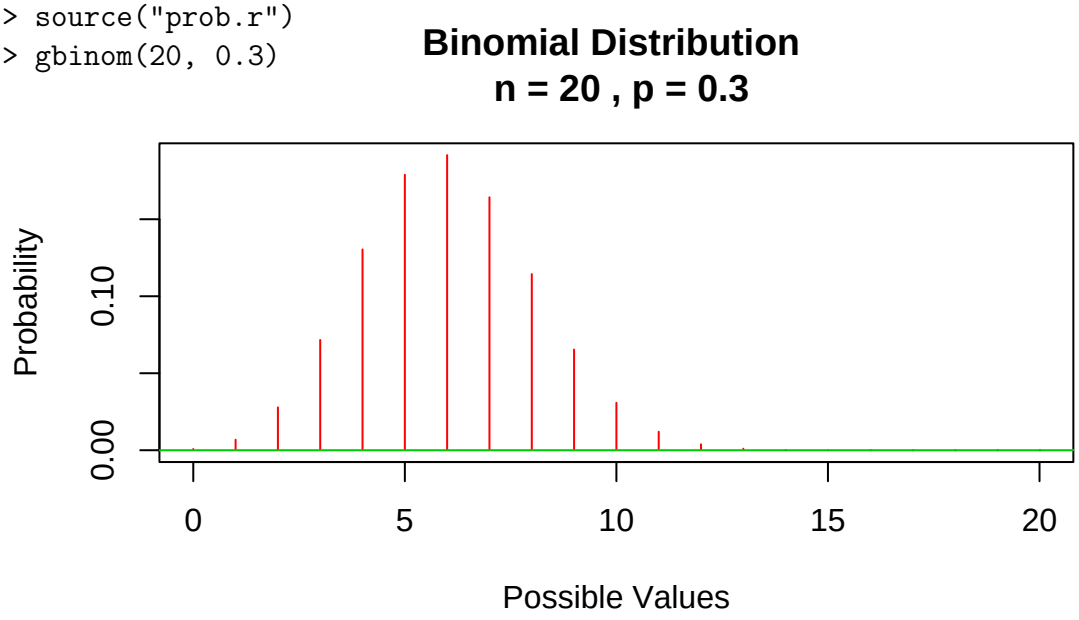
```
> xx=1:100; yy=(x-50)^2;
> plot(xx,yy,type="l")
> plot(xx,yy,type="p")
> plot(xx,yy,type="p",pch="g")
> plot(xx,yy,type="p",pch=pch=22,bg="blue",col="yellow")
> plot(xx,yy,type="n",axes=FALSE,col.lab="magenta")
> lines(xx,yy,type="h",col=5)
> title("wanna change those colors?",col.main=4)
```

`l` stands for line, `p` for points, `n` for none, `pch` for point character, `col` for color, `lab` for label and `main` for main title.

Now, as learning all about R graphics is not the point of the class, the file `prob.r` is here for you. Please download it from the R-help web page, and save it into your working directory. This file contains functions that may be used to graph and visualize the binomial, Poisson and normal distributions. This

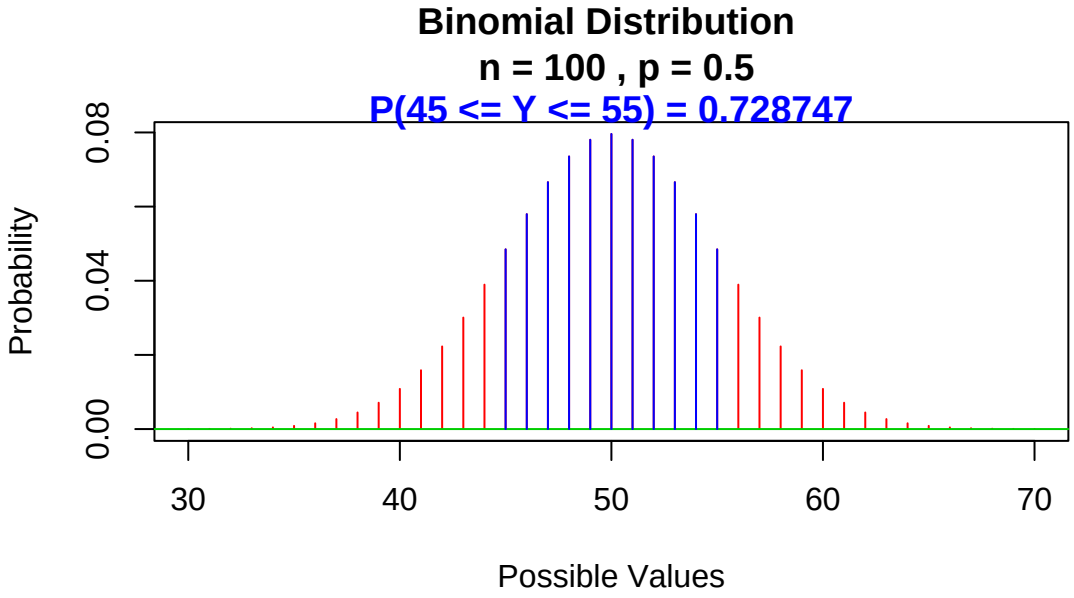
code was just made to meet your needs in this class. You need not to look at the code in it. If you'd like, of course, you can do so and re-write it if you want different colors, different titles, etc. You can either paste the text of the file, or "source" it. The next command will work if the file `prob.r` is in the working directory. Those functions are defined in the file: `gbinom`, `gpois`, and `gnorm`. Here are some examples of their use.

```
> source("prob.r")
> gbinom(20, 0.3)
```

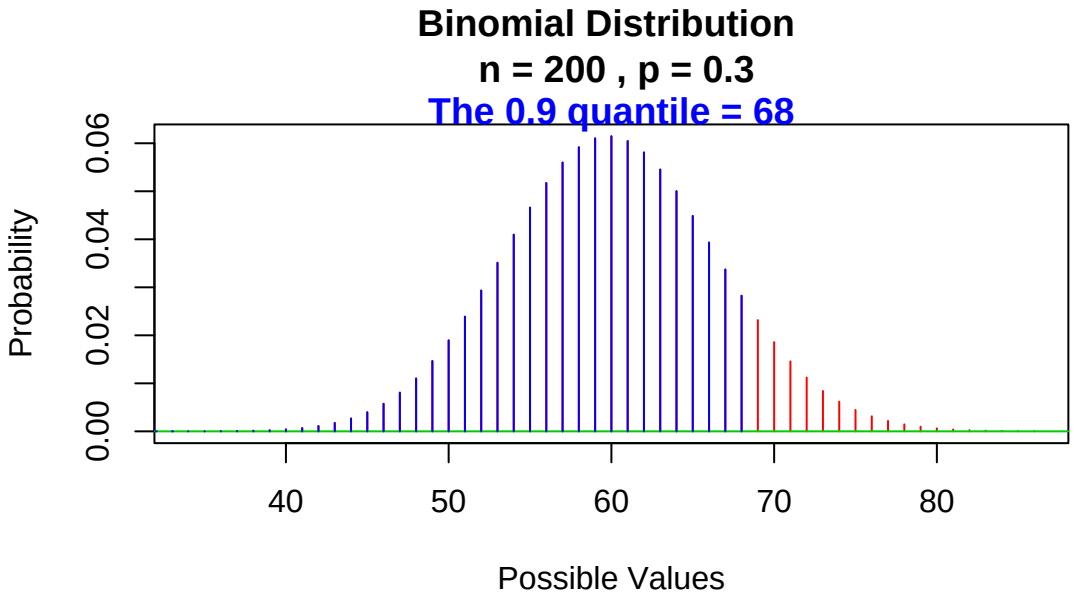


This plot will help visualize the probability of getting between 45 and 55 heads in 100 coin tosses. The next one will give you a quantile, the one we got previously with the `qbinom` function.

```
> gbinom(100, 0.5, a = 45, b = 55, scale = T)
```




```
> gbinom(200, 0.3, scale = T, quantile = 0.9)
```

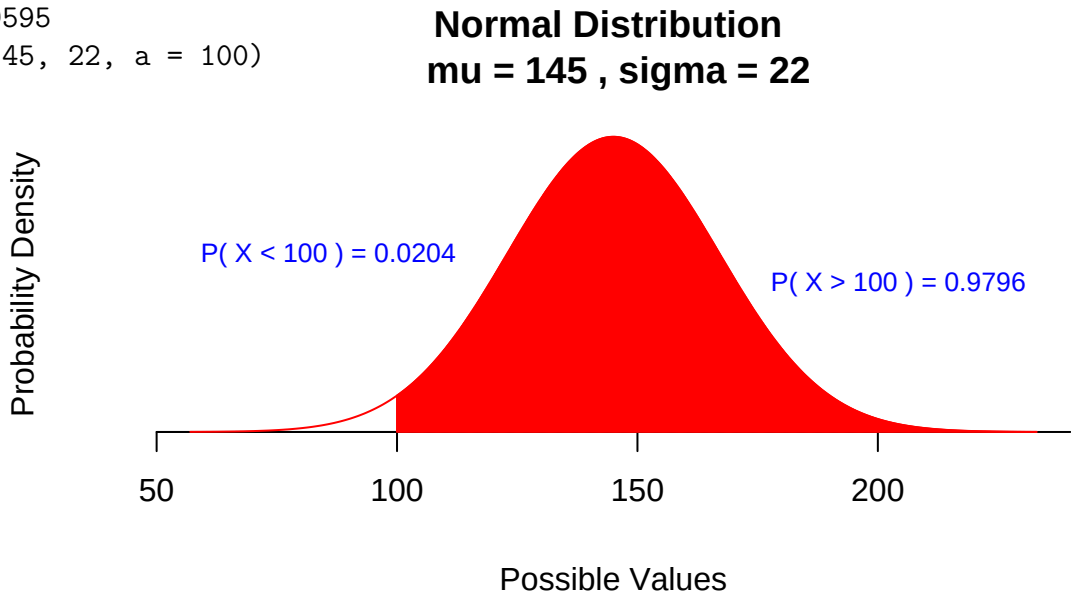


Normal Distribution Normal distributions have symmetric, bell-shaped density curves that are described by two parameters: the mean μ and the standard deviation σ. The two points of a normal density curve that are the steepest—at the "shoulders" of the curve— are precisely one standard deviation above and below the mean.

Heights of individual corn plants may be modeled as normally distributed with a mean of 145 cm and a standard deviation of 22 cm (textbook, exercise 4.29). Here are several example normal calculations using R. The commands using `gnorm` allow you to visualize the answers.

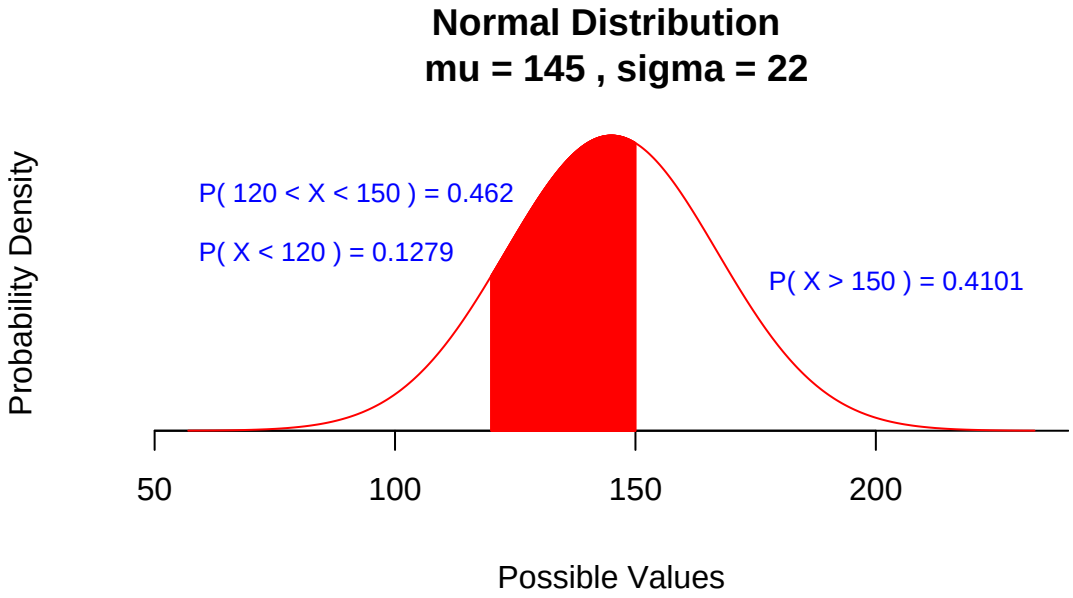
Find the proportion of plants:
. . . larger than 100cm;

```
> 1 - pnorm(100, 145, 22)
[1] 0.979595
> gnorm(145, 22, a = 100)
```



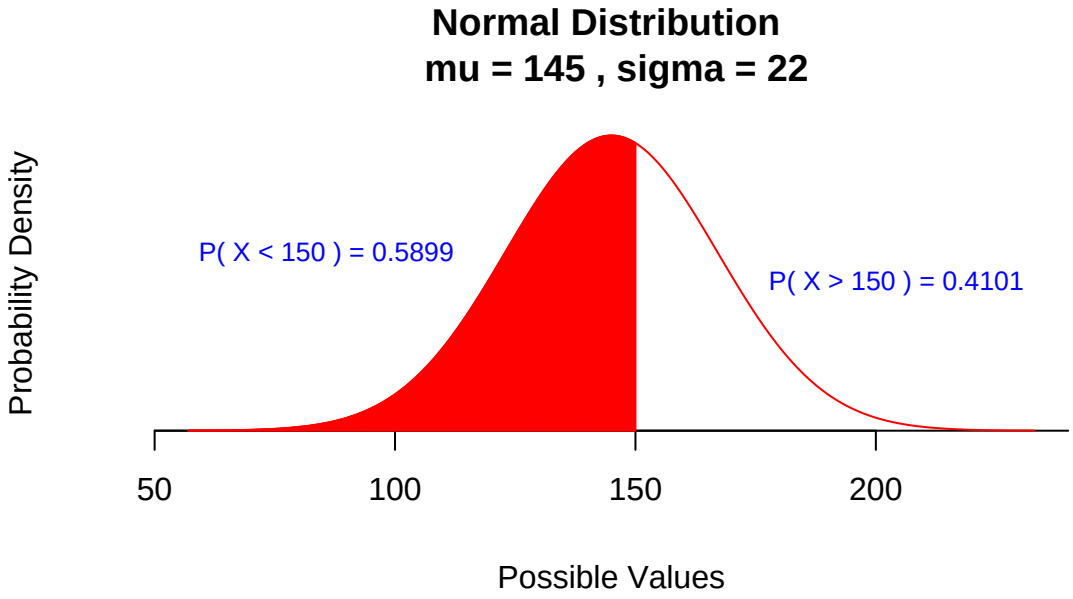
. . . between 120cm and 150cm:

```
> pnorm(150, 145, 22) - pnorm(120, 145, 22)
[1] 0.461992
> gnorm(145, 22, a = 120, b = 150)
```



. . . 150cm or less:

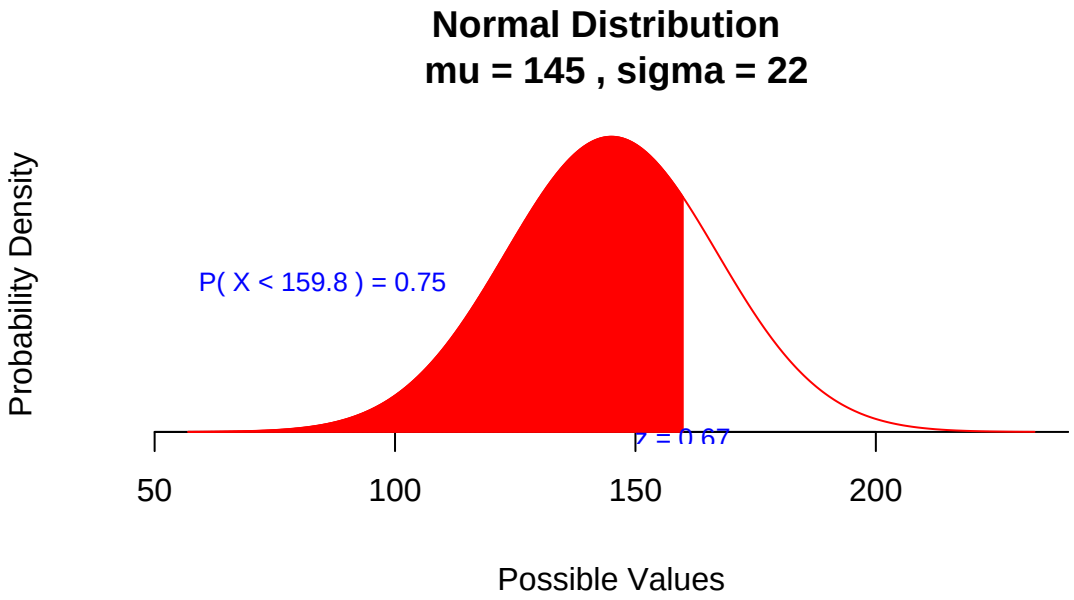
```
> pnorm(150, 145, 22)
[1] 0.5898942
> gnorm(145, 22, b = 150)
```



Find the 75th percentile.

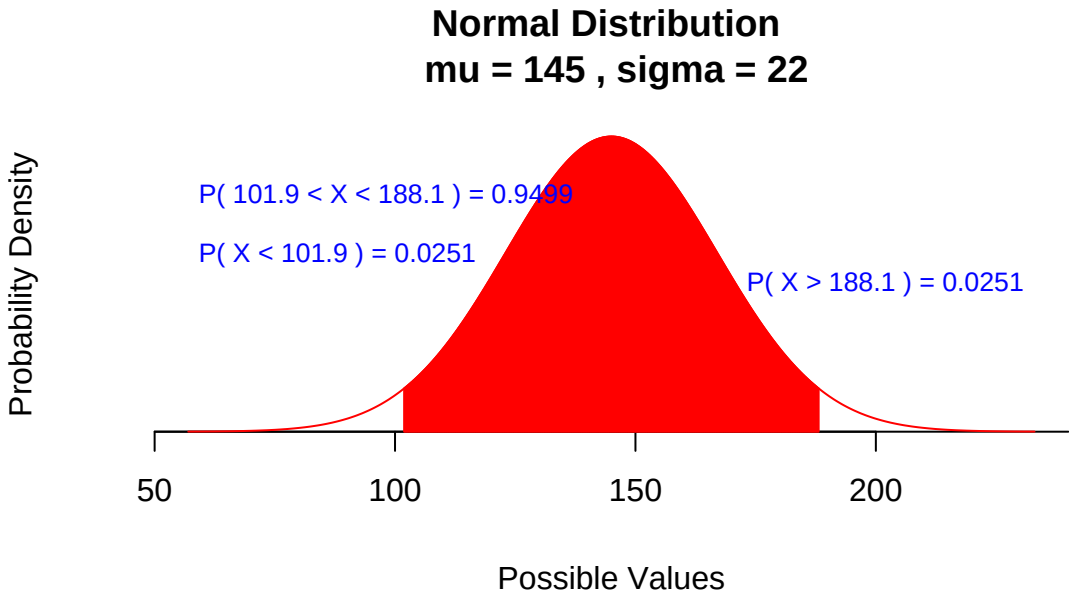
```
> qnorm(0.75, 145, 22)
```

```
[1] 159.8388
> gnorm(145, 22, quantile = 0.75)
```



Find the endpoints of middle 95% of the distribution.

```
> ab = qnorm(c(0.025, 0.975), 145, 22)
> ab
[1] 101.8808 188.1192
> gnorm(145, 22, a = round(ab[1], 1), b = round(ab[2], 1))
```



Other Distributions Other distributions work in a similar way, except that I have not yet written analogous graphing functions. Details on how to express the parameters for different probability distributions can be found from the help files. For example, to find Poisson probabilities, type `?dpois`.