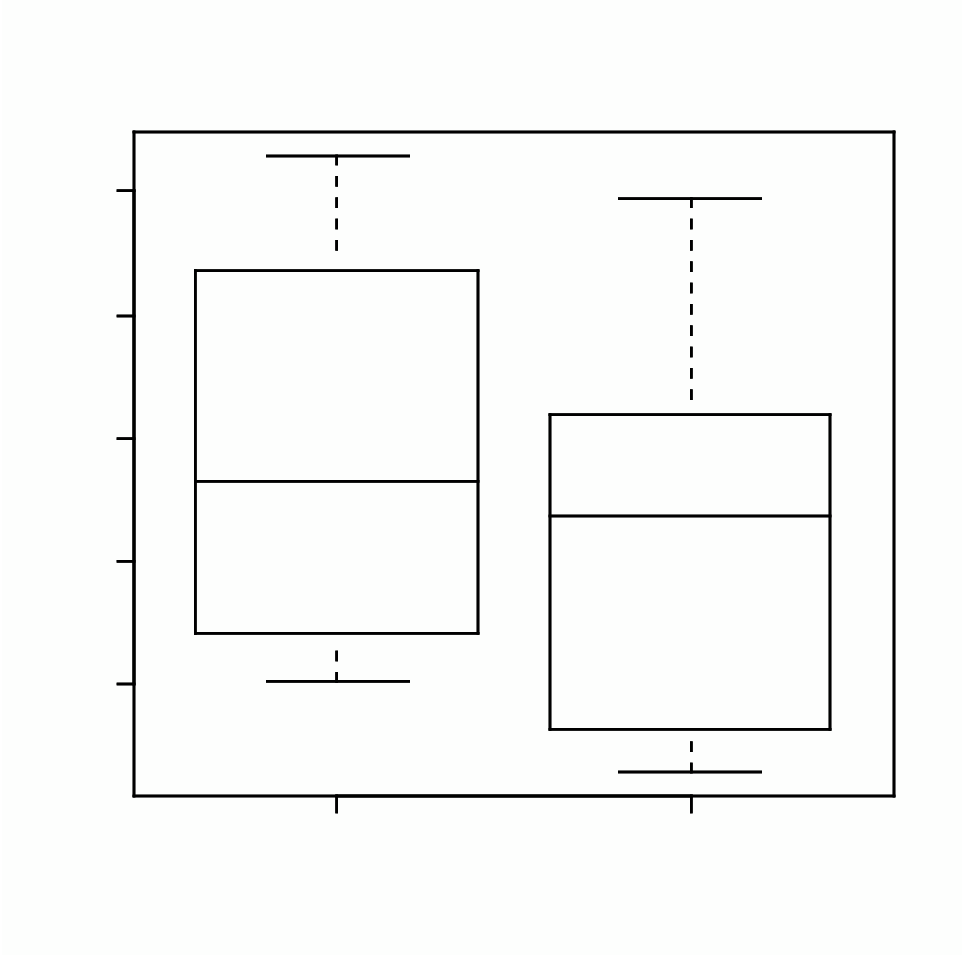


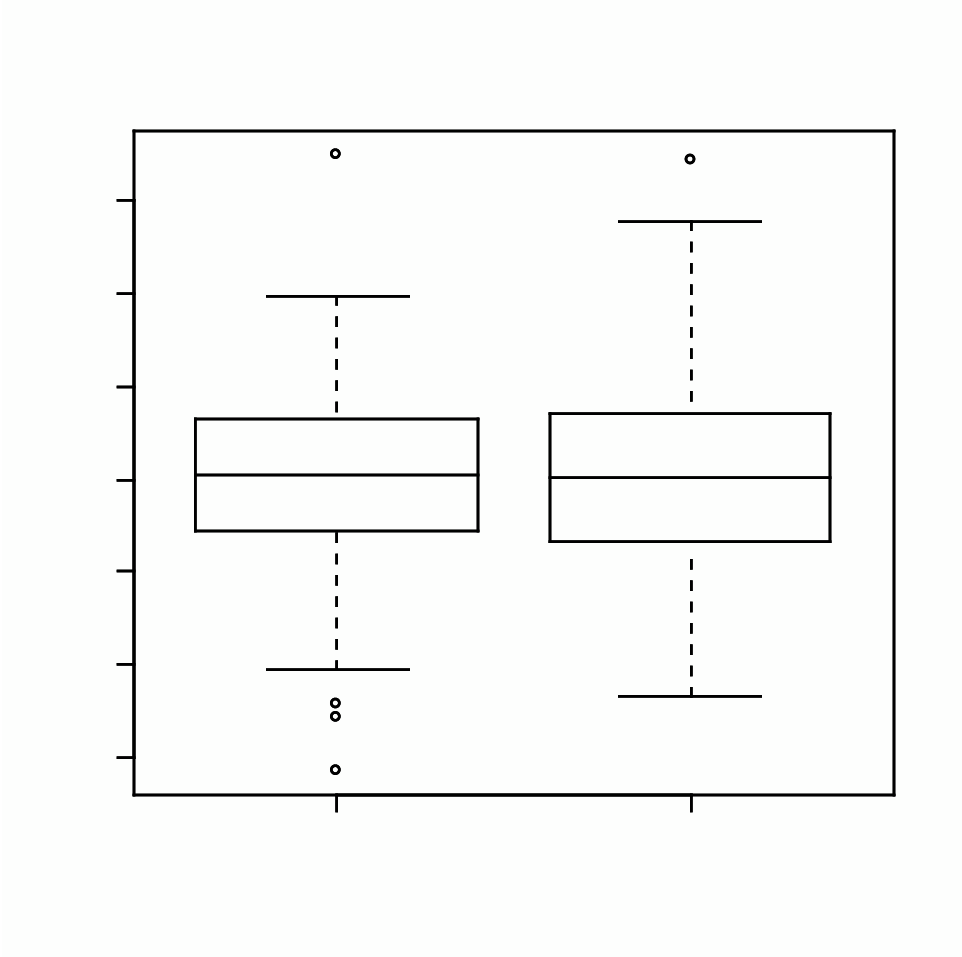
A study showing “home repair aptitude” test scores on 10 men (left) and 10 women. The y-axis is test scores”, the smallest “tick” is at 40, the next at 50, etc.

```
> summary(msample)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  45.13   47.16   53.28   53.97   59.98   66.46
> summary(fsample)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  41.40   44.44   51.77   51.12   55.76   64.68
```

Does this support the hypothesis that men know more about home repairs?

The truth is that these data are fabricated. In truth, both men and women had the same scores. But such small samples result in high variation. Here's the "population" from which they were drawn:

```
> summary(men)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  18.55   44.53   50.61   50.30   56.40   84.98
> summary(women)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  26.49   43.37   50.24   50.46   57.09   84.46
```



The samples were drawn with these commands:

```
msample <- sample(men,10)
fsample <- sample(women,10)
```

As you can see, "true" differences are slight. But random samples

show considerably more variation.

Other Simulations

Tossing a die

Simulation of 10 tosses of a die:

```
> die <- 1:6
> die
[1] 1 2 3 4 5 6
> tosses <- sample(die,10, replace=T)
> tosses
 [1] 2 1 3 5 3 1 2 4 2 6
```

The "sample" command takes a random sample (with replacement if you set "replace = T", without if you do not) of the specified size. **sample(x,n)**, for example, takes a random sample of size n from list x without replacement.

Coin Flip

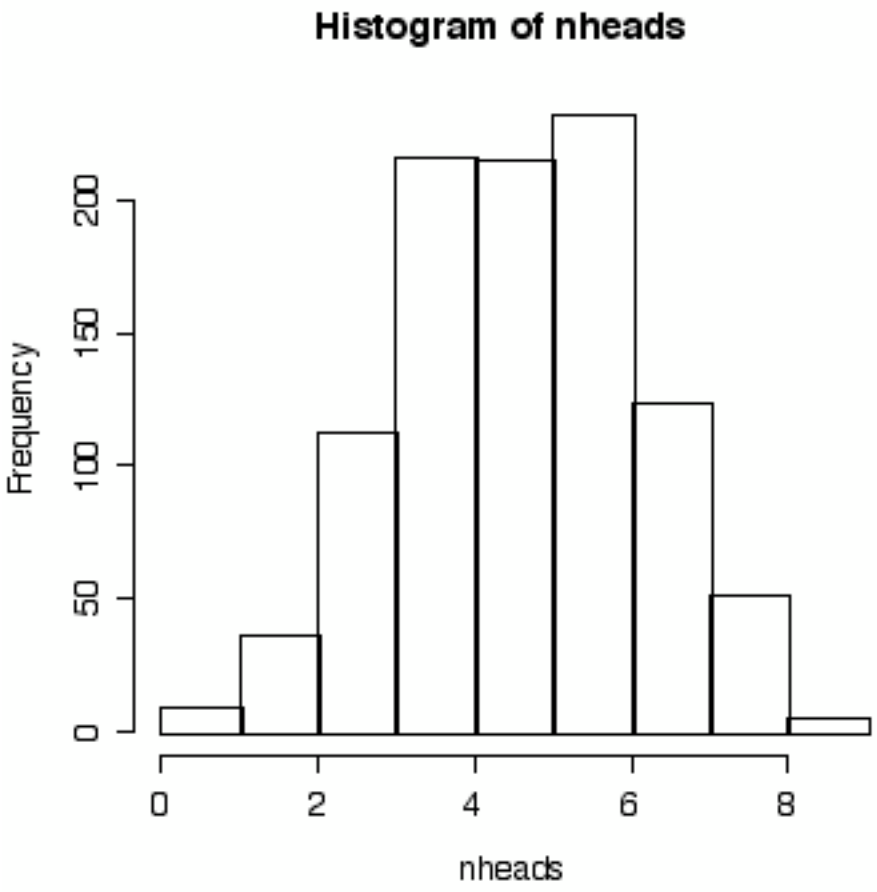
a. How many heads in 10 flips of a coin?

```
> coin <- c(0,1)
#A single flip:
> flip1 <- sample(coin,1)
> flip1
[1] 0
# Another flip:
> sample(coin,1)
[1] 0
#And another
> sample(coin,1)
[1] 1
#10 flips
> sample(coin,10,replace=T)
 [1] 0 1 0 0 0 1 1 0 1 1
# how many heads in 10 flips?  (Assuming Heads = 1)
> flip10 <- sample(coin,10,replace=T)
> sum(flip10)
```

What's the empirical probability of getting 7 heads in 10 flips of a fair coin? A poorly written bit of code that does this for you:

```
> nheads <- c()
> for (i in 1:1000){
```

```
+ flip10 <- sample(coin,10,replace=T)
+ n1 <- sum(flip10)
+ nheads <- c(n1, nheads)}
> hist(nheads)
```

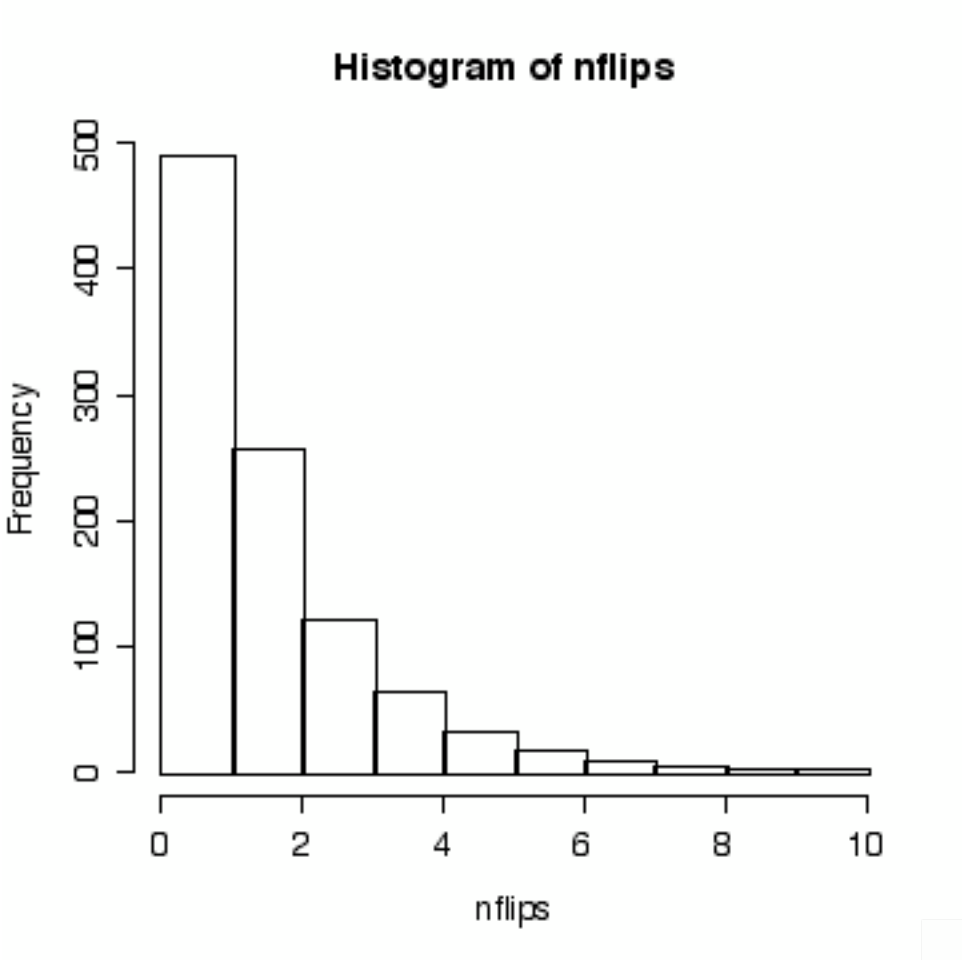



```
#Empirical Prob of 7 heads is (# of times in which 7 heads
appeared) divided by the number of trials (1000).
> sum(nheads==7)
[1] 124
> 124/1000
[1] 0.124

# If you are curious, the theoretical probability of 7 heads in 10
flips is
> dbinom(7,10,.5)
[1] 0.1171875
```

b. How many flips before the first head?

```
#This one is a little tricky.
> flip10 <- sample(coin,10,replace=T)
#Flips the coin 10 times
> flip10
 [1] 1 0 0 0 0 1 1 0 1 0
> cumsum(flip10)
 [1] 1 1 1 1 1 2 3 3 4 4
#cumsum(flip10) keeps a cumulative sum.  We want to stop when the
sum first gets to "1".
> match(1,cumsum(flip10))
[1] 1
#match identifies the first position in which a 1 appears.
Here's an empirical probability, repeating the trial 1000 times.
> nflips <- c()
> for (i in 1:1000){
+ flip20 <- sample(coin,20,replace=T)
+ firsth <- match(1,cumsum(flip20))
+ nflips <- c(firsth,nflips)}
> hist(nflips,breaks=c(0,1,2,3,4,5,6,7,8,9,10))
```



Note that approximately half of the time the first head apeared on the first flip. (Not surprisingly, really.) Approximately half again occurred on the 2nd, and half again on the third, etc.