Solutions to Exercises for lecture 1

Read Verzani-SimpleR, pages 1-8. It covers what we learned in class.

1. Create a sequence of numbers from 3 to 30 in steps of 3.

There are many ways to do this:

```
> (1:10)*3

   [1]  3  6  9 12 15 18 21 24 27 30

> (1:30)[c(F,F,T)]

   [1]  3  6  9 12 15 18 21 24 27 30

> (1:30)[1:30 %% 3 == 0]

   [1]  3  6  9 12 15 18 21 24 27 30
```

Here %% devides modulo. If a number modulo 3 is equal to 0, then it is devisible by 3. So we create a list of TRUE/FALSE depending on whether the number is divisible by 3, and then print the number from 1 to 30 that are.

```
>
```

2. Create a sequence of numbers from 1 to 30 that does not include the numbers that where mentioned above (those that are divisible by 3).

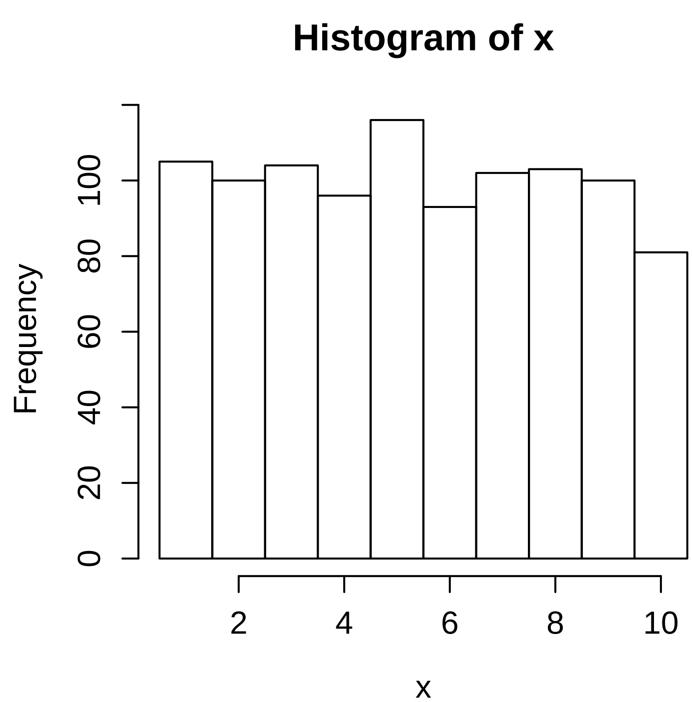
```
> (1:30)[- (1:10)*3 ]
  [1]  1  2  4  5  7  8 10 11 13 14 16 17 19 20 22 23 25 26 28 29
> (1:30)[c(T,T,F)]
  [1]  1  2  4  5  7  8 10 11 13 14 16 17 19 20 22 23 25 26 28 29
> (1:30)[1:30 %% 3 != 0]
  [1]  1  2  4  5  7  8 10 11 13 14 16 17 19 20 22 23 25 26 28 29
>
```

3. Read the documentation for the function `seq[]`, and try to use it to do 1. again.

```
> seq(3,30,by=3)
  [1]  3  6  9 12 15 18 21 24 27 30
>
```

4. Assign the variable x to be a random sample of 1000 numbers in the range 1..10, so that each occurs with equal probability. Then draw a histogram of x.

```
> x=sample(1:10,1000,rep=T)
> hist(x,breaks=0.5:10.5);v()
```

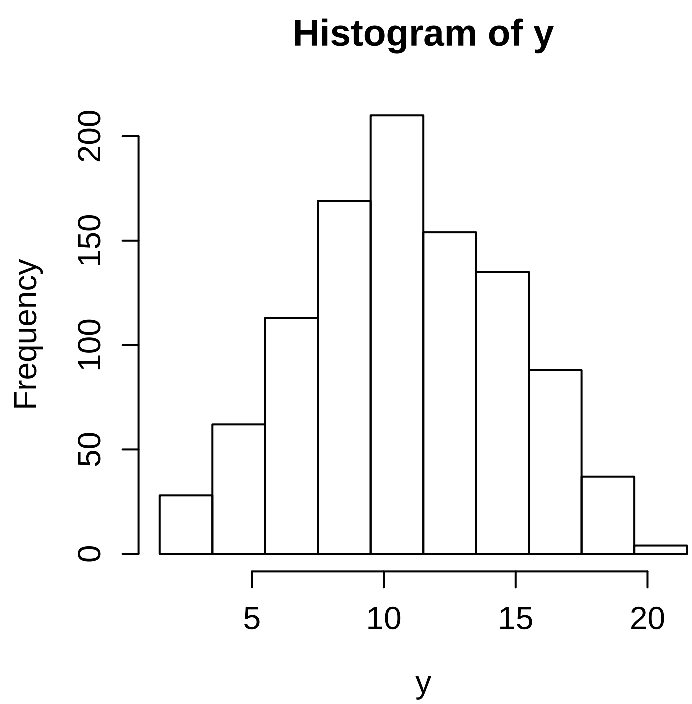


I had to tell R where the breaks should be, otherwise 1 and 2 were put in the same bin.

```
>
```

5. Now assign y to be 1000 random sums of 2 numbers, each in the range 1..10. Draw a histogram of y.

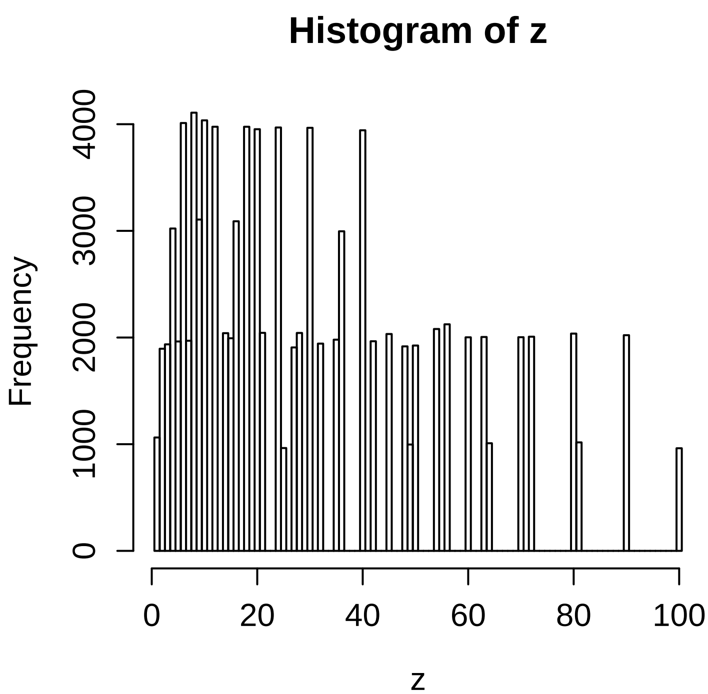
```
> y=sample(1:10,1000,rep=T)+sample(1:10,1000,rep=T)
> hist(y,breaks=seq(1.5,21.5,by=2));v()
```



```
>
```

6. Now assign z to be 10000 random products of 2 numbers, again in the same range. Again draw a histogram. Play with the parameter n, so that the result looks nice.

```
> z=sample(1:10,100000,rep=T)*sample(1:10,100000,rep=T)
> hist(z,breaks=0.5:100.5);v()
```




```
>
```

(I cheated and used 100000 numbers instead of 10000.) You can see that some numbers occur more often, because there are several ways to create them from 2 numbers between 1 and 10. The ones that have the smallest peaks, and are bigger than 10 are squares: 25, 49, 64, 81, 100. Those that have the heighest peaks can be created multiple ways: 40=5*8=4*10

Can you understand why the above histogram looks the way it does? What do the peaks mean?

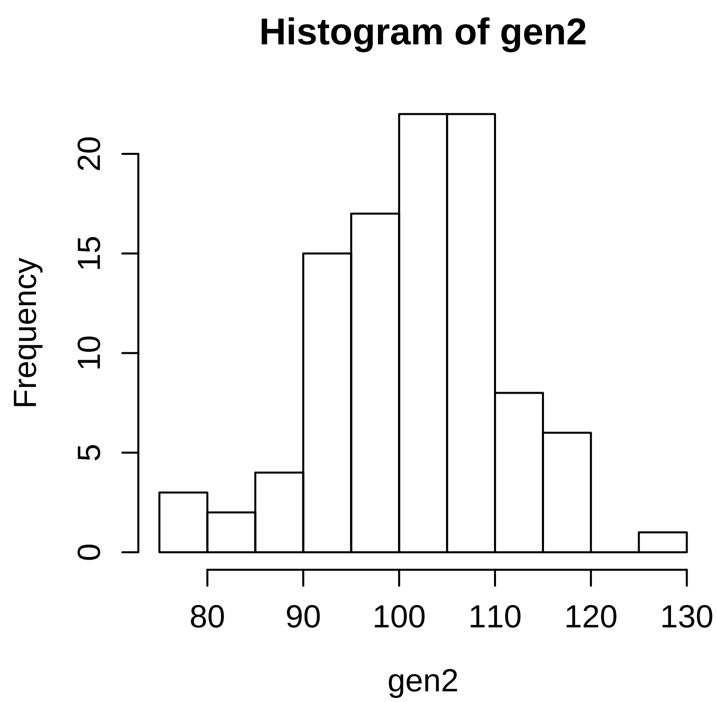
7. Now, a small genetic simulation: create a population of 100 individuals that have a genetic property that has mean 100, standard deviation 10. Set it to the variable gen1.

```
> gen1=rnorm(100,mean=100,sd=10)
>
```

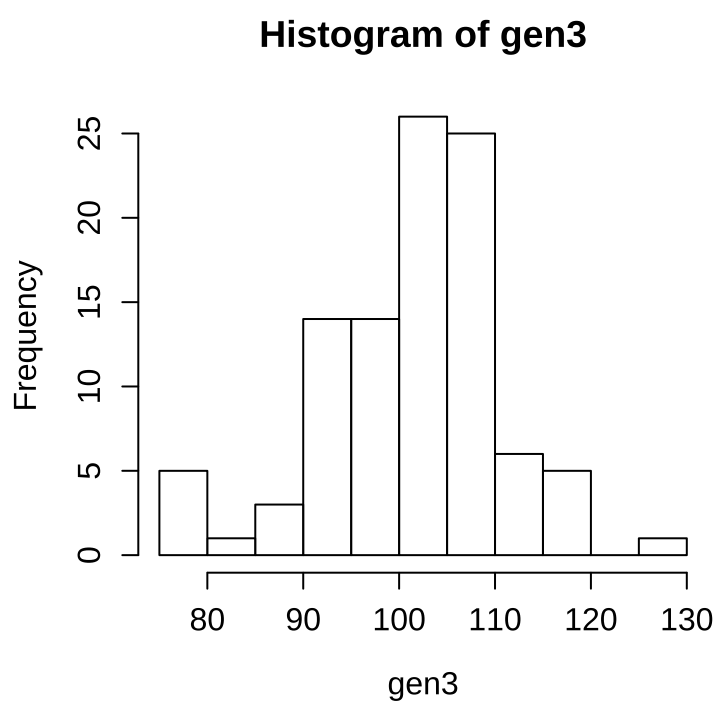
Now create the second generation: sample randomly from the above 100 individuals. This corresponds to random replication with no mutations. Assign this to gen2. Repeat as needed. Draw histograms of the distribution of the genotypes in the population as evolution proceeds.

```
> gen2=sample(gen1,100,rep=T)
```

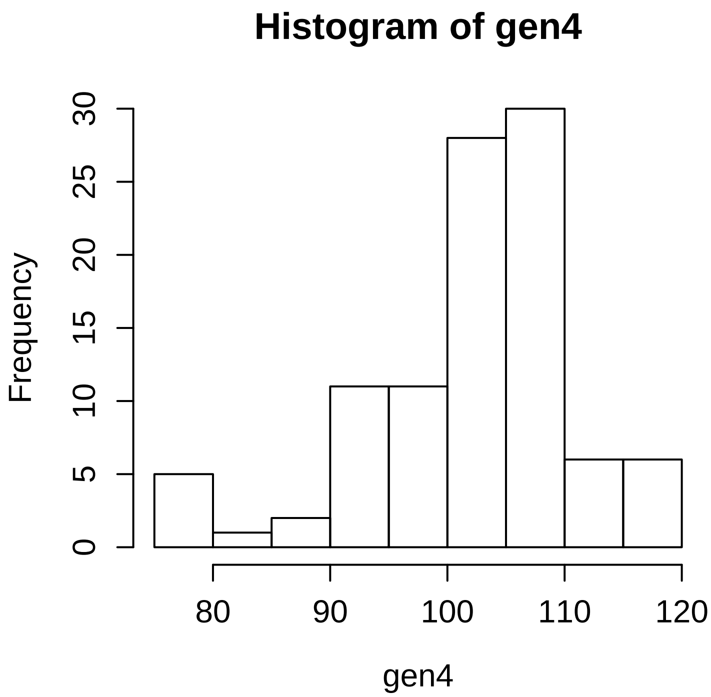
```
> hist(gen2);v()
```




```
> gen3=sample(gen2,rep=T);hist(gen3);v()
```




```
> gen4=sample(gen3,rep=T);hist(gen4);v()
```


Histogram of gen4
Frequency
0
5
10
15
20
25
30
80
90
100
110
120
gen4

>