18.05 R Tutorial: Run Length Encoding

This is a short tutorial to expand on the R reading questions. It will help you with one of the problems in pset 2.

rle(x)

```
# rle(x) stands for 'run length encoding'. It will be easiest to explain
what this means through examples. It will help with pset 2 in the
question that asks you to estimate the probability of runs in a sequence
of Bernoulli (coin flips) trials.  A run means a streak of repeats of the
same number.

# First let's make a small sequence where we can see the runs
> x = c(1,1,1,2,3,3,3,3,1,1)
# We can describe this sequence as: three 1's, then one 2, then three 3's
and two 1's.

# This is exactly what rle(x) shows us
> y = rle(x)
> y
Run Length Encoding
  lengths: int [1:4] 3 1 3 2
  values : num [1:4] 1 2 3 1
# The values vector shows the values in the order they appeared. In this
case the values of x are: 1, 2, 3, 1.

# The lengths vector shows the lenghts of the runs of each value. In this
case,
three 1's, one 2, three 3's and two 1's.

# To pick out just the lengths vector you use the syntax y$lengths
> y$lengths
[1] 3 1 3 2

# Let's look for streaks in a sequence of Bernoulli trials
# We simulate 20 Bernoulli(.5) trials using rbinon(20,1,.5).
> set.seed(1)
> y = rbinom(50,1,.5)

# y is a vector of 0's and 1's of length 20.
# We can use rle() to find the length of the longest run in y
> max(rle(y)$lengths)

# We can count the number of runs of more than 3.
> sum(rle(y)$lengths > 3)
```

```
# We can count the number of runs of exactly length 3.
> sum(rle(y)$lengths == 3)
```