R practice problems 1

Matt Boone

October 8, 2015

The key to learning R is practice, practice practice. Here are 3 problems ranked from easy to hard. In all things programming there are many ways to get to the same ending. There isn't necessarily one path to get to the answer. Take time, and don't worry if you can't figure it out, we'll post the answer next week. Just try!

Learning software is also about learning to google, so if you can't figure something out on your own, feel free to google to help. Just make sure to try on your own first!

The way this is written you should be able to easilt google your way to the medium difficulty, but might find it tougher googling your way through the hard question.

Copy and paste the data set for each problem into your own R file, and see what you can do with it!

Easy

Say you have a data set where you measured percent canopy cover at certain points. However, at a few points you could not get a value because they were actually in the middle of a pond.

```
data1<-data.frame(id=1:10, perCover=c(0.1,0.5,NA, 0.4,0.3,NA,1,0.9,0.7,0.7))
data1
```

```
##    id perCover
## 1   1      0.1
## 2   2      0.5
## 3   3       NA
## 4   4      0.4
## 5   5      0.3
## 6   6       NA
## 7   7      1.0
## 8   8      0.9
## 9   9      0.7
## 10 10      0.7
```

How would you calculate the average canopy cover?

Medium

Say you have acquired a long term data set from a banding site. Unsurprisingly, your data is riddled with NA's as there was no standard protocol across years.

```
data2<-matrix(1:50,10,5)
data2[sample.int(50,7)]<-NA
colnames(data2) <- c('id','tarsus','culmen','wingchord','tail')
data2
```

```
##       id tarsus culmen wingchord tail
##  [1,]  1     11     21        NA   41
##  [2,]  2     12     22        32   42
##  [3,]  3     13     23        33   43
##  [4,]  4     NA     24        34   44
##  [5,]  5     15     25        35   45
##  [6,]  6     16     26        36   46
##  [7,]  7     17     27        NA   NA
##  [8,] NA     18     28        38   NA
##  [9,]  9     19     29        39   NA
## [10,] 10     20     30        40   50
```

You want to take the mean of each column. How do you do this? How would you do this without taking the mean of the ID column?

Hard

Say you have a data set where you measured the temperature at different minutes. You coded time as a decimal hour (in 24 hour format). Heres your data

```
data3<-data.frame(temp=c(32,32.1,32.4,32.4,32.5,32.8,33,32.9),
                  dhour=c(22.5,22.75,22.86,22.95,22.98,23.05,23.10,23.18))

data3
```

```
##   temp dhour
## 1 32.0 22.50
## 2 32.1 22.75
## 3 32.4 22.86
## 4 32.4 22.95
## 5 32.5 22.98
## 6 32.8 23.05
## 7 33.0 23.10
## 8 32.9 23.18
```

You want to know approximately what the temperature was at 11pm (23.0 decimal hours) by picking the temperature at the closest time to 23.0.

Extra: How can you interpolate to this time point? You can use whatever interpolation method you want.