Self-Test (No Answers)
SER 50 R Workshop

S Mooney

June 20, 2017

1. Use c to create a vector named myNumbers from these numbers:
   9 9 5 8 4 2 8

2. Create a vector named myLetters of 26 lowercase letters from a to z.
   (Hint: check out the *letters* vector is built into R)

3. Combine the vectors myNumbers and myLetters columnwise into a matrix
   called my myObj.

4. Create a vector of numbers named myvec with the numbers from 19 to 10
   (i.e. 19, then 18, then 17.)

5. Sort myvec in ascending order

6. Create a logical vector called teens with the value TRUE for all numbers
   in myvec that are more than 12 (i.e. your resulting vector should start
   FALSE FALSE FALSE TRUE TRUE .)

7. Create an object containing one item named 'mynumbers' with values
   1,2,3, and 4, and one item named 'myletters' and has values e, f, and g.
   (Hint: what basic object type should this be?)

8. Run the following code:

   ```
   > data(esoph)
   > myarray <- table(esoph$agegp, esoph$alcgp, esoph$tobgp)
   ```

   (a) What value is in the myarray cell representing over age 75, consuming
       120+ g of alcohol/day and 30+ g of tobacco/day? (Hint: age 75+ is
       row 6, 120+ g is column 4, and 30+ g is depth slice 4)

   (b) How would you index myarray to return that value?

9. (a) Create a list with two items: a vector with the numbers from 1 to
       5 named 'numbers' and a character string with your name named
       'name'. Your list should look like:

```
$numbers
[1] 1 2 3 4 5

$name
[1] "Steve"
```

(Except that your name probably isn't Steve)

(b) How would you index the number 3?

10. Create a logical vector with the value TRUE for every row in the esoph data frame (that you loaded in step 1) for which ncases is greater than zero. Your vector should look like:

```
 [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE
 [8] FALSE FALSE FALSE FALSE FALSE  TRUE FALSE
[15] FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE
[22]  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE
[29] FALSE  TRUE  TRUE FALSE FALSE FALSE  TRUE
[36]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[43]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[50]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[57]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[64]  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE
[71]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[78]  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE
[85]  TRUE  TRUE  TRUE  TRUE
```

Hint: You can use the names() function to return the names of the variables in a data frame

11. create a numerical vector with the number of controls for every row in esoph for which there are any cases. Your vector should look like:

```
 [1]  1 14 23 14  3  4 46 38 21 15  7 16 14  5  4
[16]  4  4  3  4 49 22 12  6 40 21 17  6 18 15  6
[31]  4 10  7  3  6 48 14  7 34 10  9 13 12  3  1
[46]  4  2  1  1 18  6  3  5  3  1  1  1  2  1
```

12. Create a 2x2 matrix named results with 12 in the a cell, 24 in the b cell, 20 in the c cell and 8 in the d cell as follows:

```
> results <- matrix(c(12, 24, 20, 8), nrow=2, ncol=2, byrow=TRUE)
```

(a) Calculate the row and column sums using rowSums() and colSums()

(b) Calculate the row and column sums using apply()

(c) Make a 2-item list named totals with first element row sums and second element col sums

(d) Make a copy of totals named totals.copy

(e) Use mapply to add the row sums and column sums from totals to the row sums and column sums from totals.copy. Your result should look like:

```
rowSums colSums
    128     128
```

13. First, load the USArrests database built into R using data(USArrests).

   (a) What is the median number of assault arrests per 100,000 people?

   (b) What was the murder rate in the state(s) that has (have) the median rate of assault arrests have?

   (c) which states are they?
       i. Perform a linear regression predicting the number of murders by the number of assaults.
       ii. What is the slope of that regression line?
       iii. is it significantly different from 0 at $p < 0.05$?)

14. Duncan questions

   (a) Load the Duncan dataset, first by loading the car package, then loading the dataset itself:

   (b) How many of the jobs in the Duncan dataset are type=prof?

   (c) Create a new data frame named prof.jobs that is the subset of the Duncan dataset that is professional jobs.

   (d) What is the slope of the regression line predicting income from prestige among professional jobs?

15. Air quality questions

   (a) Load the airquality dataset that is built into R

   (b) Create a logical variable in the air quality dataset named niceout, which has the value true when the temperature was above 65 and below 80. (You can pick different temperatures if you prefer it to be warmer or colder out)

   (c) Use logistic regression to compute odds ratios of being nice out by month.

   (d) What are the odds of a day in June being nice compared to a day in May?

   (e) Plot the number of nice days in each month (hint: use tapply() to sum the number of nice days per month) Use either basic graphics or ggplot.