Homework Assignment #9

Matt Kline

April 26, 2017

Part 1. Higher Dimensional Arrays

1.) The built-in dataset Titanic contains information on the fate of passengers on the fatal maiden voyage of the ocean liner 'Titanic', summarized according to economic status (class), sex, age and survival. The help file (?Titanic) has more information about the data set.

a Titanic is a "table" (type class(Titanic)), but tables in R are actually arrays (type is.array(Titanic)). What are the dimensions of Titanic? What do they represent? Use dim() and dimnames().

```
dim(Titanic)

## [1] 4 2 2 2

dimnames(Titanic)

## $Class
## [1] "1st"  "2nd"  "3rd"  "Crew"
##
## $Sex
## [1] "Male"   "Female"
##
## $Age
## [1] "Child" "Adult"
##
## $Survived
## [1] "No"  "Yes"
```

b Use apply() to perform the following tasks. . Determine the total number of males and the total number of females. . Determine the total numbers of 1st, 2nd, 3rd, and Crew class passengers. . Determine how many people in each age group (children and adults) survived and how many in each age group died.

```
apply(Titanic, c("Sex"), sum)

##   Male Female
##   1731    470

apply(Titanic, c("Class"), sum)

##  1st  2nd  3rd Crew
##  325  285  706  885

apply(Titanic, c("Age", "Survived"), sum)
```

```
##        Survived
## Age       No Yes
##   Child   52  57
##   Adult 1438 654
```

c Based on the last result in Part b, does it look like the chance of surviving was better for children or for adults?

The chance of survival was better for children, roughly 25% of the adults lived while roughly 50% of the children lived.

Part 2. Input / Output

2.) Suppose you have a text file named states.txt that contains 4 lines: Alabama New York South Carolina Wyoming

After evaluating the two expressions:

```
states1 <- scan("/Users/mkline6/Downloads/Homework 9/states.txt", what = "")
states2 <- readLines("/Users/mkline6/Downloads/Homework 9/states.txt")
states1
```

```
## [1] "Alabama"  "New"      "York"     "South"    "Carolina" "Wyoming"
```

```
states2
```

```
## [1] "Alabama"        "New York"       "South Carolina" "Wyoming"
```

what is the difference between the objects states1 and states2?

The main difference is that states1 will split apart New and York while states2 will have the full string

3.)

a Use scan() to read a data file that contains numeric data. What class of object does it return?

```
class(scan("/Users/mkline6/Downloads/Homework 9/numeric_data.txt", what = ""))
```

```
## [1] "character"
```

b Use scan() to read a data file that contains character data. What class of object does it return?

```
class(scan("/Users/mkline6/Downloads/Homework 9/character_data.txt", what = ""))
```

```
## [1] "character"
```

c Use scan() to read a data file that contains a mixture of character and numeric data. What class of object does it return? Hint: Look at the help file for scan().

```
class(scan("/Users/mkline6/Downloads/Homework 9/numeric_and_character.txt", what = ""))
```

```
## [1] "character"
```

4.) Suppose you have a file, each line of which contains two numbers followed by a name. Use scan() to read the data. Turn it into a matrix with two columns, and give the matrix a dimnames attribute that uses the names in the file as row names.

```
number_name <- scan("/Users/mkline6/Downloads/Homework 9/number_name.txt", sep = " ", what = ""
matrix(number_name, nrow = 6, ncol = 2, byrow = T,
       dimnames = list(number_name[c(2, 4, 6, 8, 10, 12)], c("num", "name")))

##        num  name
## Matt   "12" "Matt"
## Joe    "52" "Joe"
## Luis   "23" "Luis"
## George "85" "George"
## Peter  "59" "Peter"
## John   "04" "John"
```

5.) Use a text editor (e.g. Notepad on Windows) to create a file consisting of the line: randomdata ¡- c(64, 38, 97, 88, 24, 14, 104, 83) Save it to a text file called randomdata.txt. Now source() the file randomdata.txt into R and confirm the randomdata vector was created

```
randomdata <- source("/Users/mkline6/Downloads/Homework 9/randomdata.txt")
randomdata

## $value
## [1]  64  38  97  88  24  14 104  83
##
## $visible
## [1] FALSE
```

6.) Create a vector called numbers which contains the values 3, 5, 8, 10, and 12. dump() the numbers to a text file called numbers.txt and delete numbers using the rm() function. Using objects() or ls(), confirm that numbers has been deleted. Now use the source() command to retrieve the vector numbers. Confirm that numbers has been recreated.

```
numbers <- c(3, 5, 8, 10)
dump("numbers", file = "/Users/mkline6/Downloads/Homework 9/numbers.txt")
rm(numbers)
objects()

## [1] "number_name" "randomdata"  "states1"     "states2"

numbers <- source("/Users/mkline6/Downloads/Homework 9/numbers.txt")
numbers

## $value
## [1]  3  5  8 10
##
## $visible
## [1] FALSE
```

7.) The built-in data set airquality contains daily air quality measurements in New York, May to September 1973.

a Write the airquality data set to a text file with write.table(). View it with a text editor (e.g. Notepad on Windows). Change the NA values to '.'s (periods) and read the changed file back into R with using read.table().

```
write.table(airquality, file = "/Users/mkline6/Downloads/Homework 9/airquality.txt")
airqualityNew <- read.table(file = "/Users/mkline6/Downloads/Homework 9/airquality.txt")
```

b What class does the variable Ozone in your new data frame belong to? What class was it in airquality?

```
class(airqualityNew$Ozone)

## [1] "integer"

class(airquality$Ozone)

## [1] "integer"
```