R Programming - Assignment #1

Myron Keith Gibert Jr

January 21, 2020

Contents

Correspondence

Please address any questions to Myron Keith Gibert Jr at mkgibertjr@msn.com. Code for this project is stored in a GitHub repository.

Introduction

For this assignment, I created two functions. The first function will create a special list vector that can store an inverted matrix based on an input matrix. The second function will create an inverted matrix using the special vector and store that matrix within that vector's environment. In addition to acquiring more practice in writing functions, this assignment involved an understanding of lexical scoping. Because of this, I will take extra steps to explain what each function does, and then perform some tests at the end to confirm that the functions work as intended.

The full instructions for this assignment can be found here.

Set Parameters

```
#set output directory? Default: outputdir <- "assignment3outputs"
outputdir <- "assignment3outputs"
#Overwrite contents of the output directory? Default: deleteoutputs <- FALSE
deleteoutputs <- TRUE
#Delete specdata/ directory after completing the analysis? Default: deletespec <- TRUE
deletehos <- TRUE
```

Debug

The debug chunk will prevent the script from running if any of the dependent variables for this analysis do not exist. This should prevent the program from erroring out after a long runtime without producing any results due to a missing variable. If modifying the input .csv and .xlsx files, it is important to leave all header information and column names intact, as the program uses this information to extract relevant data. Columns are intuitively labeled to end user convenience.

```
if (dir.exists(outputdir) && deleteoutputs == FALSE ){
  stop("Your output directory already exists!  Please delete/move
       this folder from your working directory.  Alternatively, you
       can set 'deleteoutputs' to TRUE to auto-delete this folder
       for every run. You may also choose an alternative output
       directory.")
}else{
  unlink(outputdir,recursive = TRUE)
}

if (!exists("outputdir")){
stop("outputdir variable is not defined.  Please ensure that all
     parameters in the r parameters chunk are defined.")
}

if (!exists("deleteoutputs")){
stop("deleteoutputs variable is not defined.  Please ensure that all
     parameters in the r parameters chunk are defined.")
}

if (!exists("deletehos")){
stop("deletehos variable is not defined.  Please ensure that all
     parameters in the r parameters chunk are defined.")
}

if (!dir.exists(outputdir)){dir.create(outputdir)}
```

Data

The zip file containing the data can be downloaded here:

specdata.zip 2.4MB

I have renamed the file to "ASN1_rprog_data_specdata.zip" for organization.

The zip file contains 332 comma-separated-value (CSV) files containing pollution monitoring data for fine particulate matter (PM) air pollution at 332 locations in the United States. Each file contains data from a single monitor and the ID number for each monitor is contained in the file name. For example, data for monitor 200 is contained in the file "200.csv". Each file contains three variables:

- Date: the date of the observation in YYYY-MM-DD format (year-month-day)
- sulfate: the level of sulfate PM in the air on that date (measured in micrograms per cubic meter)
- nitrate: the level of nitrate PM in the air on that date (measured in micrograms per cubic meter)

Unzipping the data

For this programming assignment I needed to unzip this file and create the directory 'specdata'. Once I unzipped the zip file, I did not make any modifications to the files in the 'specdata' directory. In each file you'll notice that there are many days where either sulfate or nitrate (or both) are missing (coded as NA). This is common with air pollution monitoring data in the United States.

```
if(
  !dir.exists("hospital-data.csv")|
   !dir.exists("outcome-of-care-measures.csv")|
   !dir.exists("Hospital_Revised_Flatfiles.pdf")
)
  {unzip("ASN3_rprog_data_hospital.zip")}
```

Plot the 30-day mortality rates for heart attack

I first needed to read the outcome data into R via the read.csv function and look at the first few rows.

```
outcome <- read.csv("outcome-of-care-measures.csv", colClasses = "character")

head(outcome)
```

I then need to make a simple histogram of the 30-day death reates from heart attack. This is column 11 in the outcome dataset.

```
# I will get a warning about NAs being introduced; that is okay
outcome[,11] <- as.numeric(outcome[,11])
```

```
## Warning: NAs introduced by coercion
```

Because I originally read the data in as character (by specifying colClasses = "character"), I need to coerce the column to be numeric. I got a warning about NAs being introduced but that is okay, as there is missing data.

Finding the best hospital in a state

Cleanup

This final command removes the unzipped "specdata" directory if the deletespec variable is set to TRUE. This reduces the overall storage burden of this project by removing the files that we no longer need access to. The zipped file remains in the working directory, so "specdata" will be recreated anyways using the command in line 93 (unzip) if it is deleted here.

```
if(deletehos== TRUE){
file.remove("hospital-data.csv")
file.remove("outcome-of-care-measures.csv")
file.remove("Hospital_Revised_Flatfiles.pdf")
}
```

```
## [1] TRUE
```