Assigmnment3

Renato Erazo

29/10/2020

## Introduction

Download the le ProgAssignment3-data.zip le containing the data for Programming Assignment 3 from the Coursera web site. Unzip the le in a directory that will serve as your working directory. When you start up R make sure to change your working directory to the directory where you unzipped the data. The data for this assignment come from the Hospital Compare web site (<http://hospitalcompare.hhs.gov>) run by the U.S. Department of Health and Human Services. The purpose of the web site is to provide data and information about the quality of care at over 4,000 Medicare-certied hospitals in the U.S. This dataset es- sentially covers all major U.S. hospitals. This dataset is used for a variety of purposes, including determining whether hospitals should be ned for not providing high quality care to patients (see <http://goo.gl/jAXFX> for some background on this particular topic). The Hospital Compare web site contains a lot of data and we will only look at a small subset for this assignment. The zip le for this assignment contains three les • outcome-of-care-measures.csv: Contains information about 30-day mortality and readmission rates for heart attacks, heart failure, and pneumonia for over 4,000 hospitals. • hospital-data.csv: Contains information about each hospital. • Hospital\_Revised\_Flatfiles.pdf: Descriptions of the variables in each le (i.e the code book). A description of the variables in each of the les is in the included PDF le named Hospital\_Revised\_Flatfiles.pdf. This document contains information about many other les that are not included with this programming assignment. You will want to focus on the variables for Number 19 (of Care Measures.csv“) and Number 11 (Data.csv”). You may nd it useful to print out this document (at least the pages for Tables 19 and 11) to have next to you while you work on this assignment. In particular, the numbers of the variables for each table indicate column indices in each table (i.e. Name" is column 2 in the outcome-of-care-measures.csv le).

# Read data

### Read the outcome data into R via the read.csv function and look at the frst few rows.

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

## Provider.Number Hospital.Name Address.1  
## 1 010001 SOUTHEAST ALABAMA MEDICAL CENTER 1108 ROSS CLARK CIRCLE  
## 2 010005 MARSHALL MEDICAL CENTER SOUTH 2505 U S HIGHWAY 431 NORTH  
## 3 010006 ELIZA COFFEE MEMORIAL HOSPITAL 205 MARENGO STREET  
## 4 010007 MIZELL MEMORIAL HOSPITAL 702 N MAIN ST  
## 5 010008 CRENSHAW COMMUNITY HOSPITAL 101 HOSPITAL CIRCLE  
## 6 010010 MARSHALL MEDICAL CENTER NORTH 8000 ALABAMA HIGHWAY 69  
## Address.2 Address.3 City State ZIP.Code County.Name Phone.Number  
## 1 DOTHAN AL 36301 HOUSTON 3347938701  
## 2 BOAZ AL 35957 MARSHALL 2565938310  
## 3 FLORENCE AL 35631 LAUDERDALE 2567688400  
## 4 OPP AL 36467 COVINGTON 3344933541  
## 5 LUVERNE AL 36049 CRENSHAW 3343353374  
## 6 GUNTERSVILLE AL 35976 MARSHALL 2565718000  
## Hospital.30.Day.Death..Mortality..Rates.from.Heart.Attack  
## 1 14.3  
## 2 18.5  
## 3 18.1  
## 4 Not Available  
## 5 Not Available  
## 6 Not Available  
## Comparison.to.U.S..Rate...Hospital.30.Day.Death..Mortality..Rates.from.Heart.Attack  
## 1 No Different than U.S. National Rate  
## 2 No Different than U.S. National Rate  
## 3 No Different than U.S. National Rate  
## 4 Number of Cases Too Small  
## 5 Number of Cases Too Small  
## 6 Number of Cases Too Small  
## Lower.Mortality.Estimate...Hospital.30.Day.Death..Mortality..Rates.from.Heart.Attack  
## 1 12.1  
## 2 14.7  
## 3 14.8  
## 4 Not Available  
## 5 Not Available  
## 6 Not Available  
## Upper.Mortality.Estimate...Hospital.30.Day.Death..Mortality..Rates.from.Heart.Attack  
## 1 17.0  
## 2 23.0  
## 3 21.8  
## 4 Not Available  
## 5 Not Available  
## 6 Not Available  
## Number.of.Patients...Hospital.30.Day.Death..Mortality..Rates.from.Heart.Attack  
## 1 666  
## 2 44  
## 3 329  
## 4 14  
## 5 9  
## 6 22  
## Footnote...Hospital.30.Day.Death..Mortality..Rates.from.Heart.Attack  
## 1   
## 2   
## 3   
## 4 number of cases is too small (fewer than 25) to reliably tell how well the hospital is performing  
## 5 number of cases is too small (fewer than 25) to reliably tell how well the hospital is performing  
## 6 number of cases is too small (fewer than 25) to reliably tell how well the hospital is performing  
## Hospital.30.Day.Death..Mortality..Rates.from.Heart.Failure  
## 1 11.4  
## 2 15.2  
## 3 11.3  
## 4 13.6  
## 5 13.8  
## 6 12.5  
## Comparison.to.U.S..Rate...Hospital.30.Day.Death..Mortality..Rates.from.Heart.Failure  
## 1 No Different than U.S. National Rate  
## 2 Worse than U.S. National Rate  
## 3 No Different than U.S. National Rate  
## 4 No Different than U.S. National Rate  
## 5 No Different than U.S. National Rate  
## 6 No Different than U.S. National Rate  
## Lower.Mortality.Estimate...Hospital.30.Day.Death..Mortality..Rates.from.Heart.Failure  
## 1 9.5  
## 2 12.2  
## 3 9.1  
## 4 10.0  
## 5 9.9  
## 6 9.9  
## Upper.Mortality.Estimate...Hospital.30.Day.Death..Mortality..Rates.from.Heart.Failure  
## 1 13.7  
## 2 18.8  
## 3 13.9  
## 4 18.2  
## 5 18.7  
## 6 15.6  
## Number.of.Patients...Hospital.30.Day.Death..Mortality..Rates.from.Heart.Failure  
## 1 741  
## 2 234  
## 3 523  
## 4 113  
## 5 53  
## 6 163  
## Footnote...Hospital.30.Day.Death..Mortality..Rates.from.Heart.Failure  
## 1   
## 2   
## 3   
## 4   
## 5   
## 6   
## Hospital.30.Day.Death..Mortality..Rates.from.Pneumonia  
## 1 10.9  
## 2 13.9  
## 3 13.4  
## 4 14.9  
## 5 15.8  
## 6 8.7  
## Comparison.to.U.S..Rate...Hospital.30.Day.Death..Mortality..Rates.from.Pneumonia  
## 1 No Different than U.S. National Rate  
## 2 No Different than U.S. National Rate  
## 3 No Different than U.S. National Rate  
## 4 No Different than U.S. National Rate  
## 5 No Different than U.S. National Rate  
## 6 Better than U.S. National Rate  
## Lower.Mortality.Estimate...Hospital.30.Day.Death..Mortality..Rates.from.Pneumonia  
## 1 8.6  
## 2 11.3  
## 3 11.2  
## 4 11.6  
## 5 11.4  
## 6 6.8  
## Upper.Mortality.Estimate...Hospital.30.Day.Death..Mortality..Rates.from.Pneumonia  
## 1 13.7  
## 2 17.0  
## 3 15.8  
## 4 19.0  
## 5 21.5  
## 6 11.0  
## Number.of.Patients...Hospital.30.Day.Death..Mortality..Rates.from.Pneumonia  
## 1 371  
## 2 372  
## 3 836  
## 4 239  
## 5 61  
## 6 315  
## Footnote...Hospital.30.Day.Death..Mortality..Rates.from.Pneumonia  
## 1   
## 2   
## 3   
## 4   
## 5   
## 6   
## Hospital.30.Day.Readmission.Rates.from.Heart.Attack  
## 1 19.0  
## 2 Not Available  
## 3 17.8  
## 4 Not Available  
## 5 Not Available  
## 6 Not Available  
## Comparison.to.U.S..Rate...Hospital.30.Day.Readmission.Rates.from.Heart.Attack  
## 1 No Different than U.S. National Rate  
## 2 Number of Cases Too Small  
## 3 No Different than U.S. National Rate  
## 4 Number of Cases Too Small  
## 5 Number of Cases Too Small  
## 6 Number of Cases Too Small  
## Lower.Readmission.Estimate...Hospital.30.Day.Readmission.Rates.from.Heart.Attack  
## 1 16.6  
## 2 Not Available  
## 3 14.9  
## 4 Not Available  
## 5 Not Available  
## 6 Not Available  
## Upper.Readmission.Estimate...Hospital.30.Day.Readmission.Rates.from.Heart.Attack  
## 1 21.7  
## 2 Not Available  
## 3 21.5  
## 4 Not Available  
## 5 Not Available  
## 6 Not Available  
## Number.of.Patients...Hospital.30.Day.Readmission.Rates.from.Heart.Attack  
## 1 728  
## 2 21  
## 3 342  
## 4 1  
## 5 4  
## 6 13  
## Footnote...Hospital.30.Day.Readmission.Rates.from.Heart.Attack  
## 1   
## 2 number of cases is too small (fewer than 25) to reliably tell how well the hospital is performing  
## 3   
## 4 number of cases is too small (fewer than 25) to reliably tell how well the hospital is performing  
## 5 number of cases is too small (fewer than 25) to reliably tell how well the hospital is performing  
## 6 number of cases is too small (fewer than 25) to reliably tell how well the hospital is performing  
## Hospital.30.Day.Readmission.Rates.from.Heart.Failure  
## 1 23.7  
## 2 22.5  
## 3 19.8  
## 4 27.1  
## 5 24.7  
## 6 23.9  
## Comparison.to.U.S..Rate...Hospital.30.Day.Readmission.Rates.from.Heart.Failure  
## 1 No Different than U.S. National Rate  
## 2 No Different than U.S. National Rate  
## 3 Better than U.S. National Rate  
## 4 No Different than U.S. National Rate  
## 5 No Different than U.S. National Rate  
## 6 No Different than U.S. National Rate  
## Lower.Readmission.Estimate...Hospital.30.Day.Readmission.Rates.from.Heart.Failure  
## 1 21.3  
## 2 19.2  
## 3 17.2  
## 4 22.4  
## 5 19.9  
## 6 20.1  
## Upper.Readmission.Estimate...Hospital.30.Day.Readmission.Rates.from.Heart.Failure  
## 1 26.5  
## 2 26.1  
## 3 22.9  
## 4 31.9  
## 5 30.2  
## 6 28.2  
## Number.of.Patients...Hospital.30.Day.Readmission.Rates.from.Heart.Failure  
## 1 891  
## 2 264  
## 3 614  
## 4 135  
## 5 59  
## 6 173  
## Footnote...Hospital.30.Day.Readmission.Rates.from.Heart.Failure  
## 1   
## 2   
## 3   
## 4   
## 5   
## 6   
## Hospital.30.Day.Readmission.Rates.from.Pneumonia  
## 1 17.1  
## 2 17.6  
## 3 16.9  
## 4 19.4  
## 5 18.0  
## 6 18.7  
## Comparison.to.U.S..Rate...Hospital.30.Day.Readmission.Rates.from.Pneumonia  
## 1 No Different than U.S. National Rate  
## 2 No Different than U.S. National Rate  
## 3 No Different than U.S. National Rate  
## 4 No Different than U.S. National Rate  
## 5 No Different than U.S. National Rate  
## 6 No Different than U.S. National Rate  
## Lower.Readmission.Estimate...Hospital.30.Day.Readmission.Rates.from.Pneumonia  
## 1 14.4  
## 2 15.0  
## 3 14.7  
## 4 15.9  
## 5 14.0  
## 6 15.7  
## Upper.Readmission.Estimate...Hospital.30.Day.Readmission.Rates.from.Pneumonia  
## 1 20.4  
## 2 20.6  
## 3 19.5  
## 4 23.2  
## 5 22.8  
## 6 22.2  
## Number.of.Patients...Hospital.30.Day.Readmission.Rates.from.Pneumonia  
## 1 400  
## 2 374  
## 3 842  
## 4 254  
## 5 56  
## 6 326  
## Footnote...Hospital.30.Day.Readmission.Rates.from.Pneumonia  
## 1   
## 2   
## 3   
## 4   
## 5   
## 6

outcome <- read.csv("outcome-of-care-measures.csv")

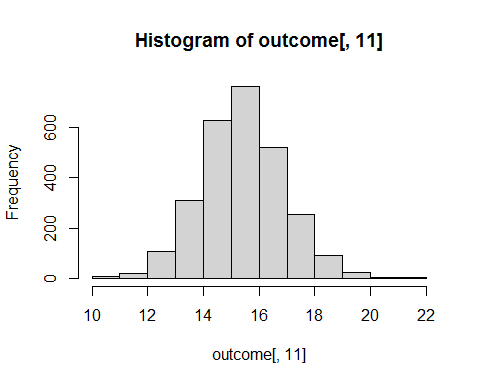
There are many columns in this dataset. You can see how many by typing ncol(outcome) (you can see the number of rows with the nrow function). In addition, you can see the names of each column by typing names(outcome) (the names are also in the PDF document. To make a simple histogram of the 30-day death rates from heart attack (column 11 in the outcome dataset), run.

Because we originally read the data in as character (by specifying colClasses = “character” we need to coerce the column to be numeric. You may get a warning about NAs being introduced but that is okay.

outcome[,11] <- as.numeric(outcome[,11])

## Warning: NAs introducidos por coerción

hist(outcome[,11])



## 2 Finding the best hospital in a state

Write a function called best that take two arguments: the 2-character abbreviated name of a state and an outcome name. The function reads the outcome-of-care-measures.csv file and returns a character vector with the name of the hospital that has the best (i.e. lowest) 30-day mortality for the specified outcome in that state. The hospital name is the name provided in the Hospital.Name variable. The outcomes can be one of attack“, failure”, or ". Hospitals that do not have data on a particular outcome should be excluded from the set of hospitals when deciding the rankings.

Handling ties. If there is a tie for the best hospital for a given outcome, then the hospital names should be sorted in alphabetical order and the frst hospital in that set should be chosen (i.e. if hospitals , , and are tied for best, then hospital should be returned).

The function should use the following template.

best <- function (state, restriccion)  
{  
   
 ## To Propercase  
 require(stringi)  
   
 restriccion <- stri\_trans\_general(restriccion, id = "Title")  
 print(restriccion)  
   
   
 ## Read outcome data  
 outcome <- read.csv("outcome-of-care-measures.csv", colClasses = "character")  
   
 ## Check that state and outcome are valid  
   
 estadoOk <- sum(sapply(outcome$State,function(x) x == state)) == 0  
   
 if (estadoOk) {  
   
 stop('Invalid state')  
 }  
   
   
 tipoRest <- c("Heart Attack", "Heart Failure", "Pneumonia")  
 siTipo <- sum(sapply(tipoRest,function(x) x == restriccion)) == 0  
   
 if(siTipo){  
   
 stop('Invalid outcome')  
 }  
   
   
   
   
 ## Return hospital name in the sate with lowest 30-day death  
   
 library(dplyr)  
 prefijo <- "Lower.Mortality.Estimate...Hospital.30.Day.Death..Mortality..Rates.from"  
indice <- sub(" ",".",restriccion)  
completo <- paste(prefijo,indice,sep = ".")  
 subOutcome <- outcome %>% select(State,Hospital.Name,completo)  
   
 ## ---- Cambio los nombre de columnas mas simple  
   
 names(subOutcome) = c("Estado","Hospital","Data")  
   
 ## --- Solo recuperar datos del estado/provincia seleccionado  
 subOutcome <- subset(subOutcome,Estado == state)  
   
   
   
   
 #texto a número  
  
 subOutcome$Data <- as.numeric(subOutcome$Data)  
   
 ## --- Eliminar NA  
   
 subOutcome <- subOutcome %>% filter(!is.na(Data))  
   
   
   
  
 ## rate  
   
   
   
   
   
 subOutcome <- subOutcome %>% arrange(Data)  
   
 #---- Iterar entre los valores menores  
   
 mejor <- subOutcome[1,3]  
   
 # --- Extraere los mejores  
   
subOutcome %>% filter(Data == mejor) %>% arrange(Data)  
   
   
   
}

## Ejemplo Best()

best("TX","heart Attack")

## Loading required package: stringi

## Warning: package 'stringi' was built under R version 4.0.3

## [1] "Heart Attack"

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

## Note: Using an external vector in selections is ambiguous.  
## i Use `all\_of(completo)` instead of `completo` to silence this message.  
## i See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.  
## This message is displayed once per session.

## Warning in best("TX", "heart Attack"): NAs introducidos por coerción

## Estado Hospital Data  
## 1 TX CYPRESS FAIRBANKS MEDICAL CENTER 9.1

## —- 3Ranking hospitals by outcome in a state

Write a function called rankhospital that takes three arguments: the 2-character abbreviated name of a state (state), an outcome (outcome), and the ranking of a hospital in that state for that outcome (num). The function reads the outcome-of-care-measures.csv le and returns a character vector with the name of the hospital that has the ranking specied by the num argument. For example, the call rankhospital(“MD”, “heart failure”, 5) would return a character vector containing the name of the hospital with the 5th lowest 30-day death rate for heart failure. The num argument can take values “, ”, or an integer indicating the ranking (smaller numbers are better). If the number given by num is larger than the number of hospitals in that state, then the function should return NA. Hospitals that do not have data on a particular outcome should be excluded from the set of hospitals when deciding the rankings. Handling ties. It may occur that multiple hospitals have the same 30-day mortality rate for a given cause of death. In those cases ties should be broken by using the hospital name. For example, in Texas ("), the hospitals with lowest 30-day mortality rate for heart failure are shown here.

rankhospital <- function(state, restriccion, num = "best") {  
 ## Read outcome data  
 ## Check that state and restriccion are valid  
   
   
   
   
   
 ## Return hospital name in that state with the given rank  
   
   
   
 ## 30-day death rate  
   
   
   
}