

Programming Assignment 3 INSTRUCTIONS: Hospital Quality

Mihai M. Craiu

3/7/2020

Introduction

Download the file Programming Assingment 3 data.zip file containing the data for Programming Assignment 3 from the Coursera web site. Unzip the file 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 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-certified hospitals in the U.S. This dataset essentially covers all major U.S. hospitals. This dataset is used for a variety of purposes, including determining whether hospitals should be fined for not providing high quality care to patients.

The Hospital Compare web site contains a lot of data and we will only look at a small subset for this assignment. The zip file for this assignment contains three files:

- 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 file (i.e the code book).

A description of the variables in each of the files is in the included PDF file named Hospital_Revised_Flatfiles.pdf. This document contains information about many other files that are not included with this programming assignment. You will want to focus on the variables for Number 19 (“Outcome of Care Measures.csv”) and Number 11 (“Hospital Data.csv”). You may find 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. “Hospital Name” is column 2 in the outcome-of-care-measures.csv file)

Detailed Instructions

Please download this document for detailed instructions about the assignment

ProgAssignment3.PDF

Data

The zip file containing the data for this assignment can be downloaded here

Programming Assingment 3 data

Data Summary

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

```
str(outcome[,c(2,7,11,17,19,23)])
```

```
## 'data.frame': 4706 obs. of 6 variables:
## $ Hospital.Name : chr "SOUTH"
## $ State : chr "AL"
## $ Hospital.30.Day.Death..Mortality..Rates.from.Heart.Attack : chr "14.3"
## $ Hospital.30.Day.Death..Mortality..Rates.from.Heart.Failure : chr "11.4"
## $ Lower.Mortality.Estimate...Hospital.30.Day.Death..Mortality..Rates.from.Heart.Failure: chr "9.5"
## $ Hospital.30.Day.Death..Mortality..Rates.from.Pneumonia : chr "10.9"
```

```
summary(outcome[,c(2,7,11,17,19,23)])
```

```
## Hospital.Name      State
## Length:4706      Length:4706
## Class :character  Class :character
## Mode :character  Mode :character
## Hospital.30.Day.Death..Mortality..Rates.from.Heart.Attack
## Length:4706
## Class :character
## Mode :character
## Hospital.30.Day.Death..Mortality..Rates.from.Heart.Failure
## Length:4706
## Class :character
## Mode :character
## Lower.Mortality.Estimate...Hospital.30.Day.Death..Mortality..Rates.from.Heart.Failure
## Length:4706
## Class :character
## Mode :character
## Hospital.30.Day.Death..Mortality..Rates.from.Pneumonia
## Length:4706
## Class :character
## Mode :character
```

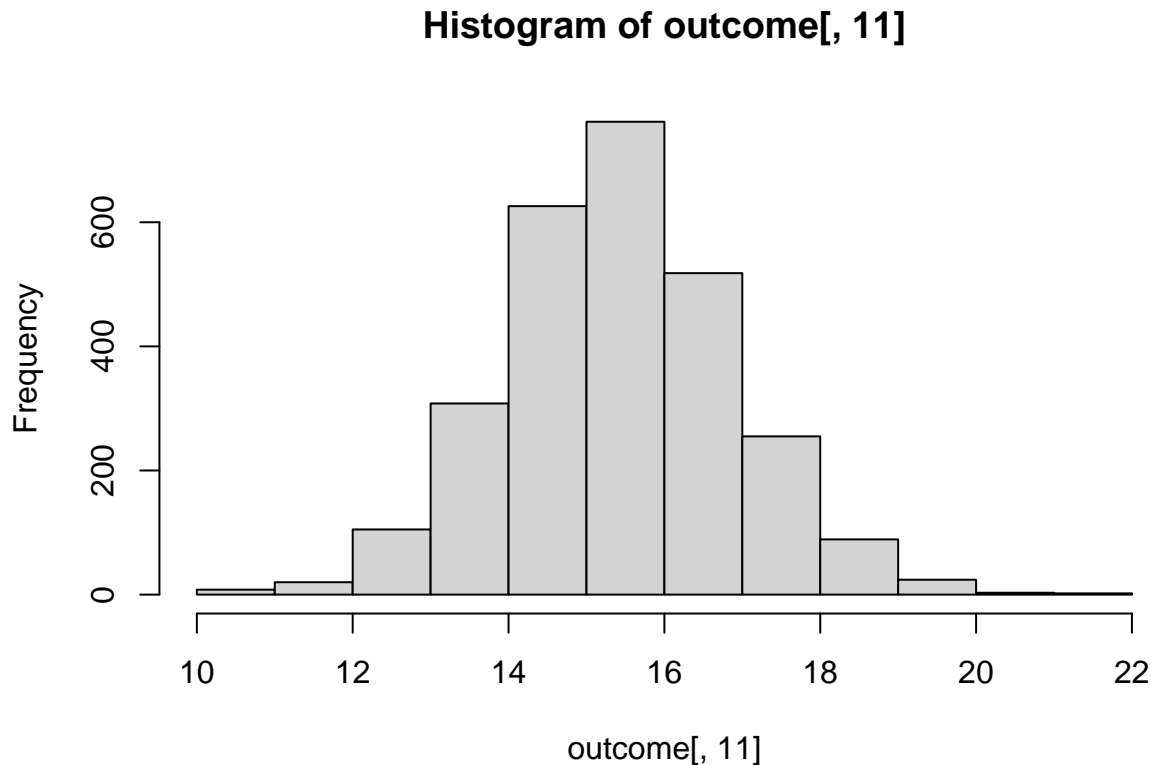
Part 1: Plot the 30-day mortality rates for heart attack

To make a simple histogram of the 30-day death rates from heart attack (column 11 in the outcome dataset), run

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

```
## Warning: NAs introducidos por coerción
```

```
## You may get a warning about NA being introduced; that is okay
hist(outcome[,11])
```



#Part 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 “heart attack”, “heart failure”, or “pneumonia”. 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 first hospital in that set should be chosen (i.e. if hospitals “b”, “c”, and “f” are tied for best, then hospital “b” should be returned).

The function should use the following template.

```
best <- function(state, outcome) {  
  
  ## Read outcome data  
  
  ## Check that state and outcome are valid  
  
  ## Return hospital name in that state with lowest 30-day death  
  ## rate  
  
}
```

The function should check the validity of its arguments. If an invalid state value is passed to `best`, the function should throw an error via the `stop` function with the exact message “invalid state”. If an invalid outcome value is passed to `best`, the function should throw an error via the `stop` function with the exact message “invalid outcome”.

Here is some sample output from the function.

```
> source("best.R")
> best("TX", "heart attack")
[1] "CYPRESS FAIRBANKS MEDICAL CENTER"
> best("TX", "heart failure")
[1] "FORT DUNCAN MEDICAL CENTER"
> best("MD", "heart attack")
[1] "JOHNS HOPKINS HOSPITAL, THE"
> best("MD", "pneumonia")
[1] "GREATER BALTIMORE MEDICAL CENTER"
> best("BB", "heart attack")
Error in best("BB", "heart attack") : invalid state
> best("NY", "hert attack")
Error in best("NY", "hert attack") : invalid outcome
```