hospitalquality

R scripts that analyze and rank hospitals based on mortality rates from the Hospital Compare data run by the U.S. Department of Health and Human Services.

Data

The data comes 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-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 (see http://goo.gl/jAXFX for some background on this particular topic).

- 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)

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

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)
```

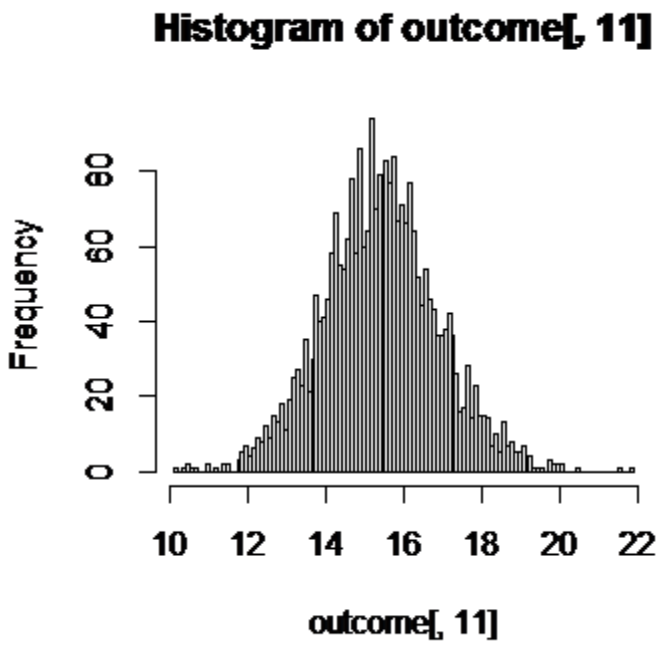
There are many columns in this dataset. You can see how many by typing ncol(outcome) (you can see the number of rows with the nrow function). In addition, you can see the names of each column by typing

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

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



---

1 Plot the 30-day mortality rates for heart attack

==========

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

```
> names(outcome)
 [1] "Provider.Number"
 [2] "Hospital.Name"
 [3] "Address.1"
 [4] "Address.2"
 [5] "Address.3"
 [6] "City"
 [7] "State"
 [8] "ZIP.Code"
 [9] "County.Name"
[10] "Phone.Number"
[11] "Hospital.30.Day.Death..Mortality..Rates.from.Heart.Attack"
[12] "Comparison.to.U.S..Rate...Hospital.30.Day.Death..Mortality..Rates.from.Heart.Attack"
[13] "Lower.Mortality.Estimate...Hospital.30.Day.Death..Mortality..Rates.from.Heart.Attack"
[14] "Upper.Mortality.Estimate...Hospital.30.Day.Death..Mortality..Rates.from.Heart.Attack"
[15] "Number.of.Patients...Hospital.30.Day.Death..Mortality..Rates.from.Heart.Attack"
[16] "Footnote...Hospital.30.Day.Death..Mortality..Rates.from.Heart.Attack"
[17] "Hospital.30.Day.Death..Mortality..Rates.from.Heart.Failure"
[18] "Comparison.to.U.S..Rate...Hospital.30.Day.Death..Mortality..Rates.from.Heart.Failure"
[19] "Lower.Mortality.Estimate...Hospital.30.Day.Death..Mortality..Rates.from.Heart.Failure"
[20] "Upper.Mortality.Estimate...Hospital.30.Day.Death..Mortality..Rates.from.Heart.Failure"
[21] "Number.of.Patients...Hospital.30.Day.Death..Mortality..Rates.from.Heart.Failure"
[22] "Footnote...Hospital.30.Day.Death..Mortality..Rates.from.Heart.Failure"
[23] "Hospital.30.Day.Death..Mortality..Rates.from.Pneumonia"
[24] "Comparison.to.U.S..Rate...Hospital.30.Day.Death..Mortality..Rates.from.Pneumonia"
[25] "Lower.Mortality.Estimate...Hospital.30.Day.Death..Mortality..Rates.from.Pneumonia"
[26] "Upper.Mortality.Estimate...Hospital.30.Day.Death..Mortality..Rates.from.Pneumonia"
[27] "Number.of.Patients...Hospital.30.Day.Death..Mortality..Rates.from.Pneumonia"
[28] "Footnote...Hospital.30.Day.Death..Mortality..Rates.from.Pneumonia"
```

[29] "Hospital.30.Day.Readmission.Rates.from.Heart.Attack"

[30] "Comparison.to.U.S..Rate...Hospital.30.Day.Readmission.Rates.from.Heart.Attack"

[31] "Lower.Readmission.Estimate...Hospital.30.Day.Readmission.Rates.from.Heart.Attack"

[32] "Upper.Readmission.Estimate...Hospital.30.Day.Readmission.Rates.from.Heart.Attack"

[33] "Number.of.Patients...Hospital.30.Day.Readmission.Rates.from.Heart.Attack"

[34] "Footnote...Hospital.30.Day.Readmission.Rates.from.Heart.Attack"

[35] "Hospital.30.Day.Readmission.Rates.from.Heart.Failure"

[36] "Comparison.to.U.S..Rate...Hospital.30.Day.Readmission.Rates.from.Heart.Failure"

[37] "Lower.Readmission.Estimate...Hospital.30.Day.Readmission.Rates.from.Heart.Failure"

[38] "Upper.Readmission.Estimate...Hospital.30.Day.Readmission.Rates.from.Heart.Failure"

[39] "Number.of.Patients...Hospital.30.Day.Readmission.Rates.from.Heart.Failure"

[40] "Footnote...Hospital.30.Day.Readmission.Rates.from.Heart.Failure"

[41] "Hospital.30.Day.Readmission.Rates.from.Pneumonia"

[42] "Comparison.to.U.S..Rate...Hospital.30.Day.Readmission.Rates.from.Pneumonia"

[43] "Lower.Readmission.Estimate...Hospital.30.Day.Readmission.Rates.from.Pneumonia"

[44] "Upper.Readmission.Estimate...Hospital.30.Day.Readmission.Rates.from.Pneumonia"

[45] "Number.of.Patients...Hospital.30.Day.Readmission.Rates.from.Pneumonia"

[46] "Footnote...Hospital.30.Day.Readmission.Rates.from.Pneumonia"

Plot the 30-day mortality rates for heart attack

==========

```
> outcome[, 11] <- as.numeric(outcome[, 11])
> hist(outcome[, 11], breaks = 100)
```

hist(outcome[,11]

---

2 Finding the best hospital in a state:

==========

File name: best.R

==========

```
best <- function(state, outcome) {

  ## Read the outcome data
  dat <- read.csv("outcome-of-care-measures.csv", colClasses = "character")
  ## Check that state and outcome are valid
  if (!state %in% unique(dat[, 7])) {
    stop("invalid state")
  }
  switch(outcome, `heart attack` = {
    col = 11
  }, `heart failure` = {
    col = 17
  }, pneumonia = {
    col = 23
  }, stop("invalid outcome"))
  ## Return hospital name in that state with lowest 30-day death rate
  df = dat[dat$State == state, c(2, col)]
  df[which.min(df[, 2]), 1]
}
> best("TX", "heart attack")
[1] "CYPRESS FAIRBANKS MEDICAL CENTER"
Warning message:
In which.min(df[, 2]) : NAs introduced by coercion
```

```
> best("TX", "heart failure")
[1] "FORT DUNCAN MEDICAL CENTER"
Warning message:
In which.min(df[, 2]) : NAs introduced by coercion
> best("MD", "heart attack")
[1] "JOHNS HOPKINS HOSPITAL, THE"
Warning message:
In which.min(df[, 2]) : NAs introduced by coercion
>  best("MD", "pneumonia")
[1] "GREATER BALTIMORE MEDICAL CENTER"
> > best("BB", "heart attack")
Error: unexpected '>' in ">"
```

---

rankhospital.R.

==========

3-Ranking hospitals by outcome in a state

==========

```
rankhospital <- function(state, outcome, num = "best") {

   ## Read the outcome data
   dat <- read.csv("outcome-of-care-measures.csv", colClasses = "character")
   ## Check that state and outcome are valid
   if (!state %in% unique(dat[, 7])) {
      stop("invalid state")
   }
```

```
    switch(outcome, `heart attack` = {
        col = 11
    }, `heart failure` = {
        col = 17
    }, pneumonia = {
        col = 23
    }, stop("invalid outcome"))
    dat[, col] = as.numeric(dat[, col])
    df = dat[dat[, 7] == state, c(2, col)]
    df = na.omit(df)
    nhospital = nrow(df)
    switch(num, best = {
        num = 1
    }, worst = {
        num = nhospital
    })
    if (num > nhospital) {
        return(NA)
    }
    ## Return hospital name in that state with the given rank 30-day death rate

    o = order(df[, 2], df[, 1])
    df[o, ][num, 1]
}
> rankhospital("TX", "heart failure", 4)
[1] "DETAR HOSPITAL NAVARRO"
Warning message:
In rankhospital("TX", "heart failure", 4) : NAs introduced by coercion
>
```

```
> rankhospital("MD", "heart attack", "worst")
[1] "HARFORD MEMORIAL HOSPITAL"
Warning message:
In rankhospital("MD", "heart attack", "worst") : NAs introduced by coercion
>
>
> rankhospital("MN", "heart attack", 5000)
[1] NA
Warning message:
In rankhospital("MN", "heart attack", 5000) : NAs introduced by coercion
```

---

rankall.R.

==========

4 Ranking hospitals in all states

==========

```
rankall <- function(outcome, num = "best") {
   ## Read the outcome data
   dat <- read.csv("outcome-of-care-measures.csv", colClasses = "character")
   ## Check that state and outcome are valid
   states = unique(dat[, 7])
   switch(outcome, `heart attack` = {
      col = 11
   }, `heart failure` = {
```

```
    col = 17
}, pneumonia = {
    col = 23
}, stop("invalid outcome"))

## Return hospital name in that state with the given rank 30-day death rate
dat[, col] = as.numeric(dat[, col])
dat = dat[, c(2, 7, col)]  # leave only name, state, and death rate
dat = na.omit(dat)
# head(dat) Hospital.Name State 1 SOUTHEAST ALABAMA MEDICAL CENTER AL 2
# MARSHALL MEDICAL CENTER SOUTH AL 3 ELIZA COFFEE MEMORIAL HOSPITAL AL 7 ST
# VINCENT'S EAST AL 8 DEKALB REGIONAL MEDICAL CENTER AL 9 SHELBY BAPTIST
# MEDICAL CENTER AL
# Hospital.30.Day.Death..Mortality..Rates.from.Heart.Attack 1 14.3 2 18.5 3
# 18.1 7 17.7 8 18.0 9 15.9
rank_in_state <- function(state) {
    df = dat[dat[, 2] == state, ]
    nhospital = nrow(df)
    switch(num, best = {
        num = 1
    }, worst = {
        num = nhospital
    })
    if (num > nhospital) {
        result = NA
    }
    o = order(df[, 3], df[, 1])
    result = df[o, ][num, 1]
    c(result, state)
```

```
  }
  output = do.call(rbind, lapply(states, rank_in_state))
  output = output[order(output[, 2]), ]
  rownames(output) = output[, 2]
  colnames(output) = c("hospital", "state")
  data.frame(output)
}
head(rankall("heart attack", 20), 10)
                       hospital state
AK                         <NA>    AK
AL     D W MCMILLAN MEMORIAL HOSPITAL    AL
AR   ARKANSAS METHODIST MEDICAL CENTER    AR
AZ JOHN C LINCOLN DEER VALLEY HOSPITAL    AZ
CA              SHERMAN OAKS HOSPITAL    CA
CO           SKY RIDGE MEDICAL CENTER    CO
CT           MIDSTATE MEDICAL CENTER    CT
DC                         <NA>    DC
DE                         <NA>    DE
FL     SOUTH FLORIDA BAPTIST HOSPITAL    FL
Warning message:
In rankall("heart attack", 20) : NAs introduced by coercion
-----------------------------------------------------
> tail(rankall("pneumonia", "worst"), 3)
                              hospital state
WI MAYO CLINIC HEALTH SYSTEM - NORTHLAND, INC    WI
WV                   PLATEAU MEDICAL CENTER    WV
WY          NORTH BIG HORN HOSPITAL DISTRICT    WY
Warning message:
In rankall("pneumonia", "worst") : NAs introduced by coercion
```

```
> tail(rankall("heart failure"), 10)
                                                hospital state
TN                    WELLMONT HAWKINS COUNTY MEMORIAL HOSPITAL    TN
TX                                FORT DUNCAN MEDICAL CENTER    TX
UT VA SALT LAKE CITY HEALTHCARE - GEORGE E. WAHLEN VA MEDICAL CENTER    UT
VA                                SENTARA POTOMAC HOSPITAL    VA
VI                     GOV JUAN F LUIS HOSPITAL & MEDICAL CTR    VI
VT                                  SPRINGFIELD HOSPITAL    VT
WA                                HARBORVIEW MEDICAL CENTER    WA
WI                              AURORA ST LUKES MEDICAL CENTER    WI
WV                                FAIRMONT GENERAL HOSPITAL    WV
WY                                CHEYENNE VA MEDICAL CENTER    WY
Warning message:
In rankall("heart failure") : NAs introduced by coercion
```
