Problem Set 3

Martin Aragoneses

July 23, 2015

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, providing 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. There are 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 (codebook).

This assignment will focus on the variables for Number 19 ("Outcome of Care Measures.csv") and Number 11 ("Hospital Data.csv"). 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).

Housekeeping

First we change the working directory to be the same as where the data for this assignment is:

```
setwd("~/Desktop/data/github/Data-Science-R/Coursera/2-R-Programming")
```

Data Exploration

Read the outcome data into R

```
outcome <- read.csv("outcome-of-care-measures.csv", colClasses = "character")
#  read the data in as character :colClasses = "character"
head(outcome[, 1:11])
```

```
##   Provider.Number                     Hospital.Name
## 1          010001 SOUTHEAST ALABAMA MEDICAL CENTER
## 2          010005    MARSHALL MEDICAL CENTER SOUTH
## 3          010006   ELIZA COFFEE MEMORIAL HOSPITAL
## 4          010007         MIZELL MEMORIAL HOSPITAL
## 5          010008      CRENSHAW COMMUNITY HOSPITAL
## 6          010010    MARSHALL MEDICAL CENTER NORTH
##                     Address.1 Address.2 Address.3         City State
## 1     1108 ROSS CLARK CIRCLE                            DOTHAN    AL
## 2 2505 U S HIGHWAY 431 NORTH                              BOAZ    AL
## 3         205 MARENGO STREET                           FLORENCE    AL
## 4              702 N MAIN ST                                OPP    AL
## 5        101 HOSPITAL CIRCLE                            LUVERNE    AL
```

```
## 6    8000 ALABAMA HIGHWAY 69                      GUNTERSVILLE    AL
##   ZIP.Code County.Name Phone.Number
## 1    36301     HOUSTON   3347938701
## 2    35957    MARSHALL   2565938310
## 3    35631  LAUDERDALE   2567688400
## 4    36467   COVINGTON   3344933541
## 5    36049    CRENSHAW   3343353374
## 6    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
```

```
ncol(outcome)  # number of variables
```

```
## [1] 46
```

```
nrow(outcome)  #number of observations
```

```
## [1] 4706
```

```
names(outcome)[1:11] # vector with 11 first variable names
```

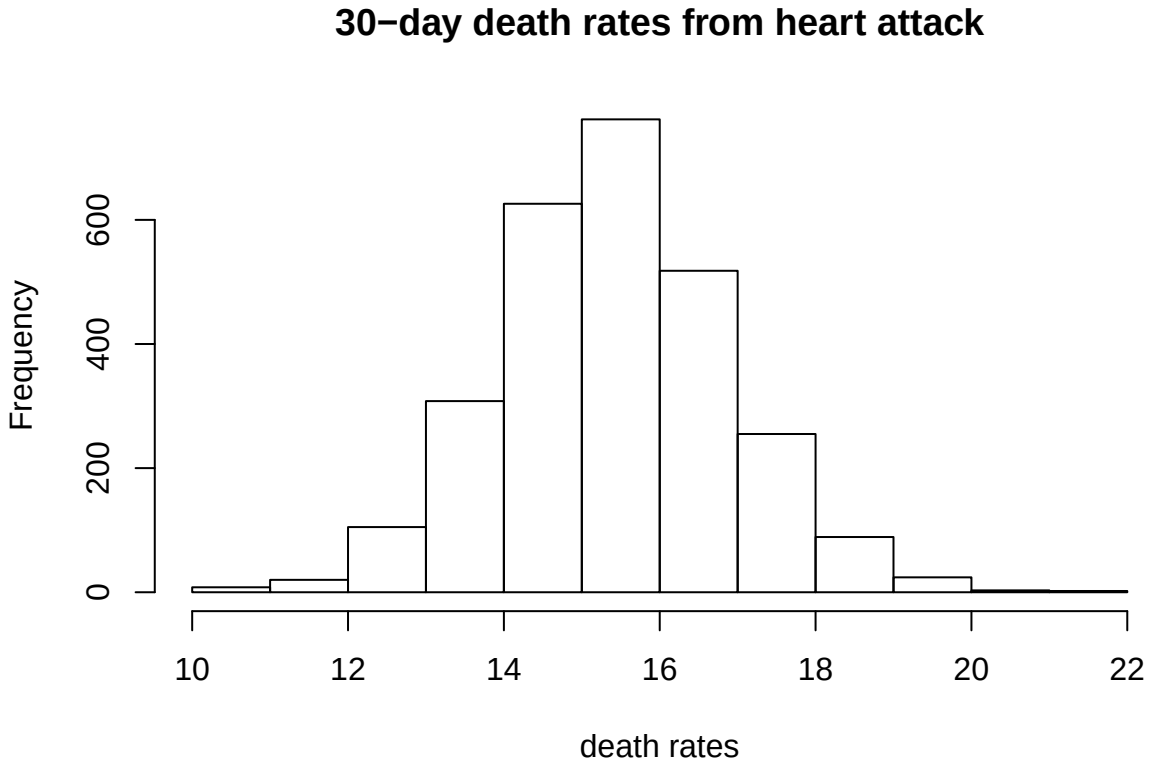
```
##  [1] "Provider.Number"
##  [2] "Hospital.Name"
##  [3] "Address.1"
##  [4] "Address.2"
##  [5] "Address.3"
##  [6] "City"
##  [7] "State"
##  [8] "ZIP.Code"
##  [9] "County.Name"
## [10] "Phone.Number"
## [11] "Hospital.30.Day.Death..Mortality..Rates.from.Heart.Attack"
```

Histogram of the 30-day death rates from heart attack

```
outcome[, 11] <- as.numeric(outcome[, 11]) # we need to coerce the "character" column to be numeric.
```

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

```
hist(outcome[, 11], main ="30-day death rates from heart attack", xlab = "death rates")
```


30−day death rates from heart attack
Frequency
0
200
400
600
10
12
14
16
18
20
22
death rates

Analysis