Hospital Data Report

Zach Christensen

September 4, 2015

Healthcare-Associated Infections, State

The Healthcare-Associated Infections (HAI) measures - state data. These measures are developed by Centers for Disease Control and Prevention (CDC) and collected through the National Healthcare Safety Network (NHSN). They provide information on infections that occur while the patient is in the hospital. These infections can be related to devices, such as central lines and urinary catheters, or spread from patient to patient after contact with an infected person or surface. Many healthcare associated infections can be prevented when the hospitals use CDC-recommended infection control steps.

The metric reported is standardized infection ratios (SIRs). SIRs compare the actual number of Healthcare-Assiocated Infections at each hospital to the predicted number of infections. These ratios are adjusted for various risk factors and population traits. More information about SIRs and how they are calculated can be found here: http://www.leapfroggroup.org/media/file/SIRCalc.pdf. .

Import Libraries and Data Set

The analysis relies on several open source libraries, and uses data directly from the healthdata.gov API.

```r
# Load Required Libraries
library(dplyr)
library(tidyr)
library(reshape2)
library(ggplot2)

# Load data from healthdata.gov
states <- read.csv("http://data.medicare.gov/api/views/k2ze-bqvw/rows.csv?accessType=DOWNLOAD")
```

Preprocessing and Exploration

The dataset containes 7: State, Measure.Name, Measure.ID, Score, Footnote, Measure.Start.Date, Measure.End.Date. Before the analysis, the data can be restructured and some redundant information removed.

```r
# Date Range
startDate <- unique(states$Measure.Start.Date)
endDate <- unique(states$Measure.End.Date)

# Split the Measure.ID column
states <- states %>% mutate("Measure" = substring(text = Measure.ID, first = 1, last = 5),
                            "Type" = substring(text = Measure.ID, first = 7,
                                               last = length(Measure.ID)))

# List of different measures
measures <- unique(filter(states, Type == "SIR")[,c("Measure", "Measure.Name")])

# Then only select needed columns
```

```r
states <- subset(states, select = c("State", "Measure.Name", "Measure", "Type", "Score"))
states <- dcast(states, State + Measure ~ Type)

# Join the measures so we get measure names with data
states <- left_join(states, measures)

# Remove rows which have NA
states <- states[-which(is.na(states$SIR)),]

# Preview Data
head(states[,1:5])
```

```
##   State Measure CI_LOWER CI_UPPER   SIR
## 1    AK   HAI_1    0.276    0.913 0.526
## 2    AK   HAI_2    0.848    1.965 1.319
## 3    AK   HAI_3    0.443    1.230 0.763
## 4    AK   HAI_4    0.270    2.049 0.849
## 5    AK   HAI_5    0.129    0.978 0.405
## 6    AK   HAI_6    0.662    0.968 0.804
```

The data for this set was collected from 10/01/2013 to 09/30/2014, with information about 6 different measures collected from 53 states or regions. The 6 measures are various healthcare associated infections:

```
## [1] Catheter-Associated Urinary Tract Infections (CAUTI)
## [2] Clostridium difficile (C.diff.) Laboratory-identified Events (Intestinal infections)
## [3] Methicillin-resistant Staphylococcus Aureus (MRSA) Blood Laboratory-identified Events (Bloodstrea
## [4] Central line-associated blood stream infections (CLABSI)
## [5] Surgical Site Infection from abdominal hysterectomy (SSI: Hysterectomy)
## [6] Surgical Site Infection from colon surgery (SSI: Colon)
## 18 Levels: C.diff Lower Confidence Limit ... Surgical Site Infection from colon surgery (SSI: Colon)
```
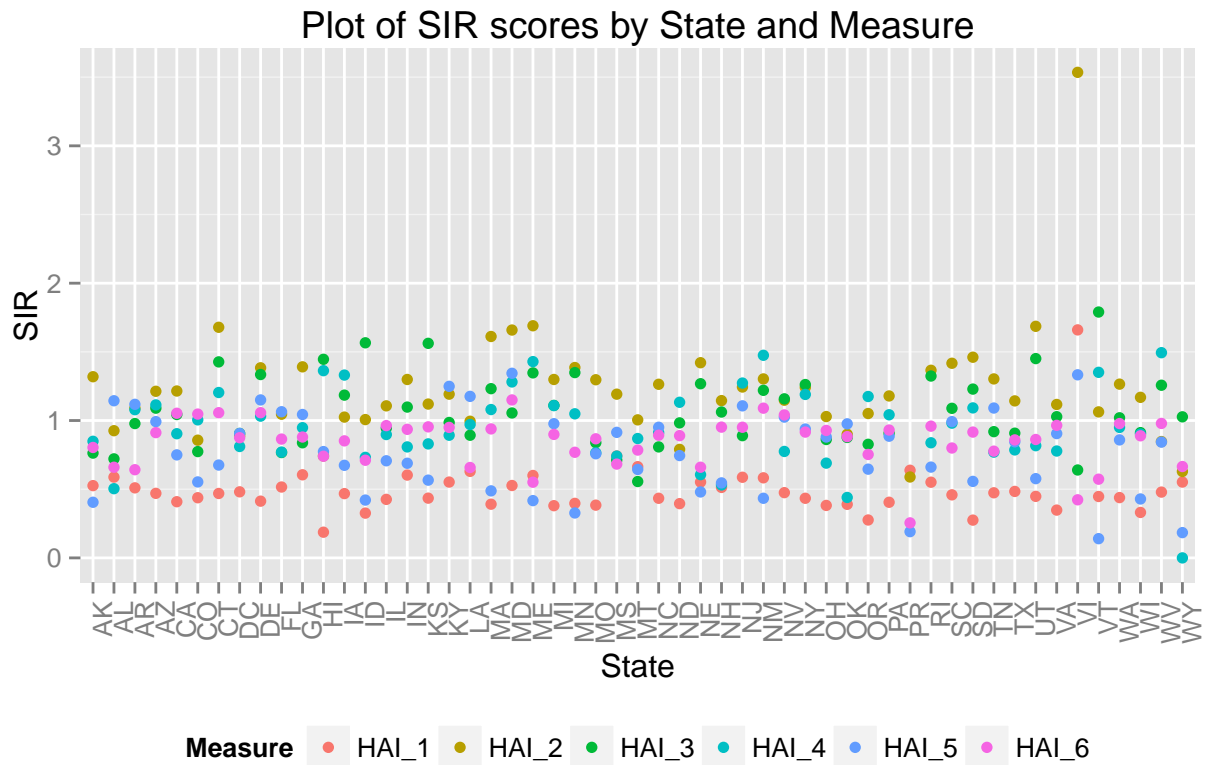
Analysis

We can plot the SIR scores for each Measure by state.

```r
# Plot of Scores by state and Measure
qplot(State, SIR, data = states, col = Measure, main = "Plot of SIR scores by State and Measure") + ther
```

2

Plot of SIR scores by State and Measure

Visually, it looks like the the SIR scores for HAI_2, Catheter-Associated Urinary Tract Infections (CAUTI) are usually the highest, while HAI_1, Central line-associated blood stream infections (CLABSI) are the lowest. Below the average score is grouped by measure.

```
## Source: local data frame [6 x 2]
##
##   Measure mean(SIR)
## 1   HAI_1 0.4933585
## 2   HAI_2 1.2253019
## 3   HAI_3 1.0623462
## 4   HAI_4 0.9484314
## 5   HAI_5 0.7763774
## 6   HAI_6 0.8499434
```