Homework 3

Data Science I

The data we will be using for this assignment is from data.cms.gov accessed at https://data.cms.gov/Medicare-Inpatient/Inpatient-Prospective-Payment-System-IPPS-Provider/97k6-zzx3. The data include hospital-specific charges for the more than 3,000 U.S. hospitals that receive Medicare Inpatient Prospective Payment System (IPPS) payments for Fiscal Year (FY) 2011. The data are for the top 100 most frequently billed discharges as categorized by Medicare Severity Diagnosis Related Group (MS-DRG). These DRGs represent more than 7 million discharges or 60 percent of total Medicare IPPS discharges. Before you begin this assignment do a little bit of research and familiarize yourself with DRG codes (https://en.wikipedia.org/wiki/Diagnosis-related_group).

1. Read in the data.

2. Make a single plot with boxplots of the average medicare payments by DRG code. This is the amount of money that medicare pays for the DRG code. Note that this may take some finessing to make a plot that has relevant information and is visually appealing! What do you notice from the plots?

3. Do the same as (2) for the average total payments. This is the total amount paid for the claim (including the part that is paid for by the patient). What do you notice from the plots? Does the plot differ from the plot you made in part 2?

4. Read pages 98 to 103 of Modern Data Science with R. The functions spread() and gather() are from the package tidyr and are very useful for manipulating your data from narrow to wide format and back again! Select the variables 'DRG.Definition', 'Provider.Id', 'Provider.State' , and 'Average.Medicare.Payments' and use the spread() function to put the data into wide format with a column for each of the DRG codes containing the average medicare payments.

5. Write a for loop to calculate the mean, median, and standard deviation of the average medicare payments for each DRG code column. Which DRG code has the max mean, median, and sd and what are these values? The minimum mean, median, and sd and what are these values? Remember that if you are performing any operation more than twice you should write a function to do it!

6. Do the same with a single map and a single apply statement.

7. Use a do() statement to return the provider in each state with the most expensive average medicare payments for the DRG code '870 - SEPTICEMIA OR SEVERE SEPSIS W MV 96+ HOURS'. Join with the original data to find the name of the most expensive provider in New York. Do the same for two other DRG codes of your choosing. Remember the rule for copying and pasting code more than 2 times!