**Business Questions -----along with a list of specific statements/hypothesis with reasoning/intuition**

The average overall ratings of private hospitals are higher rating than public ones. The higher rating reflects better performances that could be further mapped into statistics of hospital service measured by time and effective care, readmission and death, efficiency, et.al.

According to the [2014 American Hospital Association Annual Survey](http://www.aha.org/research/rc/stat-studies/fast-facts.shtml), there are 5,686 hospitals in the United States, with 2,904 public hospitals and 1,060 private. There are a total of 795,603 staffed beds in public hospitals and 118,910 staffed beds in private hospitals. Public hospitals had about 33.6 million admissions annually while private hospitals had about 1.8 million admissions annually. Although the total public hospital almost triple the amount of private, they are serving 18 times more patients. Therefore, it is reasonable to expect the private hospital to provide better serving

**Data Sources and Description**

For hypothesis 1, the hospital performance datasets, the datasets are from Medicare Hospital Compare website (<https://data.medicare.gov/data/hospital-compare>), provided by the Centers for Medicare & Medicaid Services. These data allow you to compare the quality of care at over 4,000 Medicare-certified hospitals across the country. The pros of this data source is it doesn’t require extra cleaning, and it record the specific measurement related to different treatment or conditions in numerical values. The challenges are that there are millions of data points at observational levels, and measurements were recorded towards different conditions and treatments, described in the data dictionary of 86 pages. These requires broad and specific domain knowledge of health care service as well as the programming ability to handle big data.

**Data Preprocessing Steps (so far)**

The datasets from the databases of Medicare web are quite a complex. It integrates every single detail of performance measurements from service, equipment, hospital setting to procedures, regulation rules/policy and different payment programs. Each domain involves new terms and require certain level of understanding towards each part. While I have taught myself most of the terms and policies and built some degree of understanding towards all the data, I am still working on to improve my understanding towards measurement details. Majority of measurement data are recorded at the observational levels, these will help me to sort and find the related the data to perform the analysis

Other preprocessing for the following preliminary analysis includes:

Check the data quality, data size, data type, data levels and missing values

Combine different variables for analysis, simplify names of variables and levels.

Change "Not Available" value to NA value to perform statistical analysis

**Specific Approach/Analysis (so far)**

Compare different hospital rating by ownerships and different owners with mean, sd, bar chart and pie chart

Drill down (one level) to find the difference among owners, this leads to more interesting results which will be further explored.

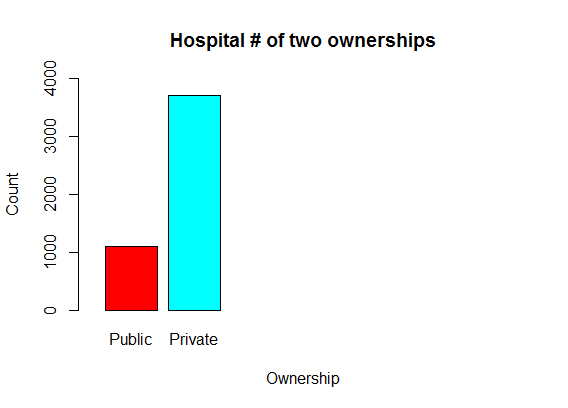
**Specific Outcomes/Insights and Summary and Conclusions**

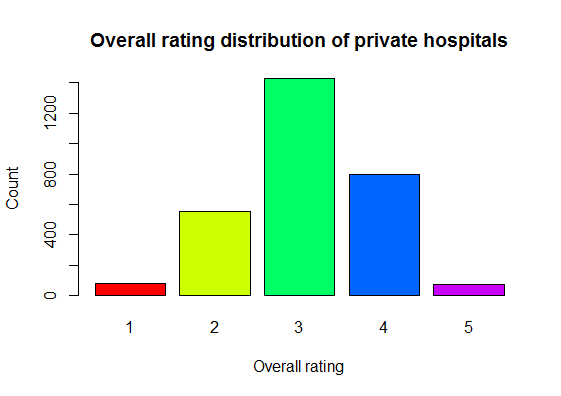
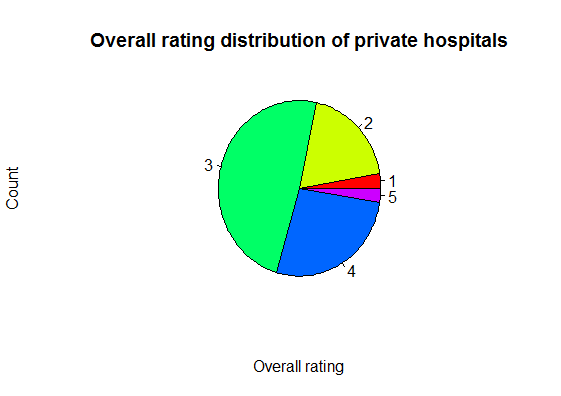
The amount of private hospitals most triple the public hospitals. If I just look at the overall rating roughly by the ownership, there is not much difference. But the rating distribution of these two types displays more. More than 50% of the public hospitals have a medium rating 3, with a total higher rating (4 and 5) less than a quarter and a total lower rating (1 and 2) accounting for about 25%. For private hospitals, the total higher rating accounts over 25% and more than the total lower rating, despite that they also have about half rating at the middle level . Considering the rating samples of private hospitals are 4-5 times over the public hospitals, this result implies that the private hospitals provide better service overall.

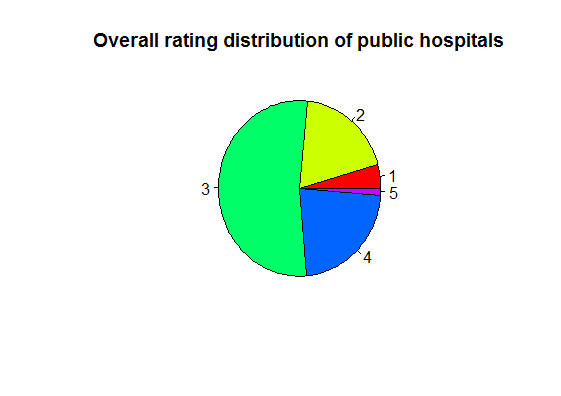
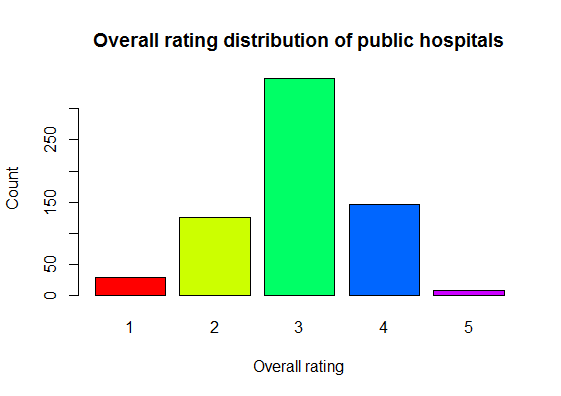
Ownership Avg.rating Sd

1 Public 2.969559 0.7995723

2 Private 3.078579 0.8146654





****

While drilling down from different ownerships to different types of owners, the majority of public hospitals are from Hospital District and Local, and the majority of private hospitals are from non-profit private and proprietary. The federal state hospital has the lowest rating and the physician owned hospitals have the highest average rating. In general, the most private owned hospitals have higher rating than the government owned except for the proprietary type, which is interesting and need to further looked into.

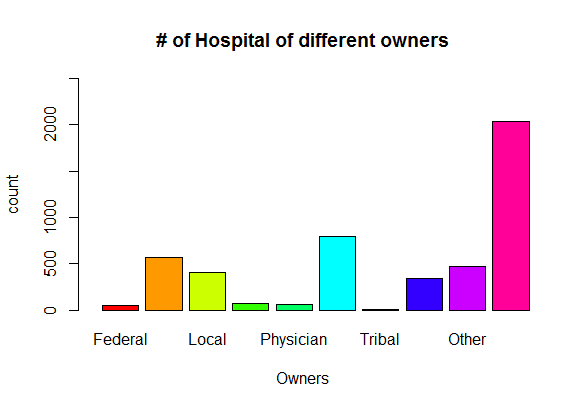
Count of hospitals of different owners

Federal District Local State Physician Proprietary

46 566 406 67 64 796

Tribal Church Other Private

8 344 473 2037

****

Owners Avg.rating Ownership

1 Federal 2.941176 Public

2 District 3.028736 Public

3 Local 2.954733 Public

4 State 2.638298 Public

5 Physician 3.888889 Private

6 Proprietary 2.845659 Private

7 Tribal 2.500000 Public

8 Church 3.186495 Private

9 Other 3.157068 Private

10 Private 3.120452 Private

**Key learning (so far)**

Check the data quality and data types.

The functions used to structure data vector and data frame, and to manipulate the data types

Obtain the key statistics parameters and statistical analysis and plotting.

Self-teaching the geographical mapping.

**Specific Next Steps**

Further mapping the performance difference onto each measurement, there are 15-20 measurement and each of them have 5-20 indexes.

Geographically mapping the performance rating across the country.

**References (so far)**

<http://www.npinstitute.com/public-vs-private-hospitals-s/1852.htm>

<https://data.medicare.gov/data/hospital-compare>