Hw1 IEMS 308

SHENGJIE LIU

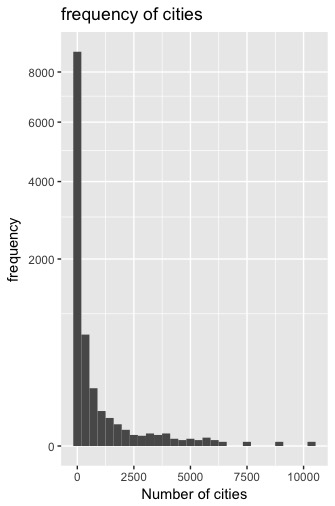
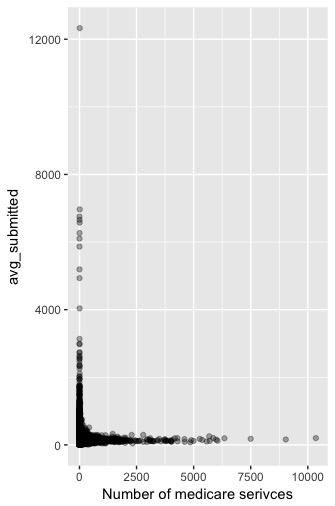
**Executive Summary**:

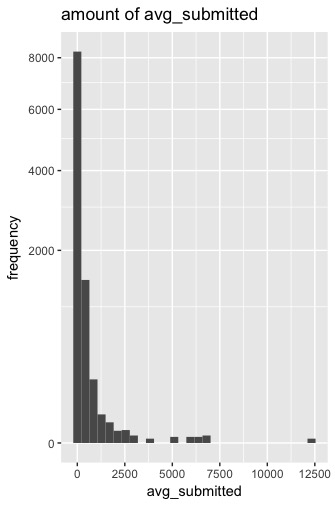
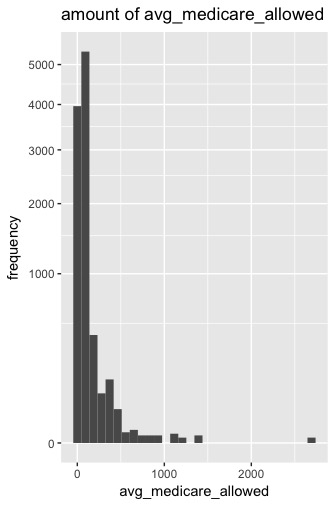
Based on the medicare data description, we can know that this data source provides many features of the medicare provider. To fully utilize this data source, we should take advantage of data and info provided by both parties. For example, we can analyze the medicare cover rate based on clustering analysis.

Clustering analysis is the task of grouping a set of objects in such a way that objects in the samegroup are more similar to each other than to those in other groups. In this case, medicare coverrate = medicare allowed amount / medicare submitted amount. By using clustering analysis, wecan analyze relationship between medicare cover rate and the number of medicare services provided based on grouping cities.

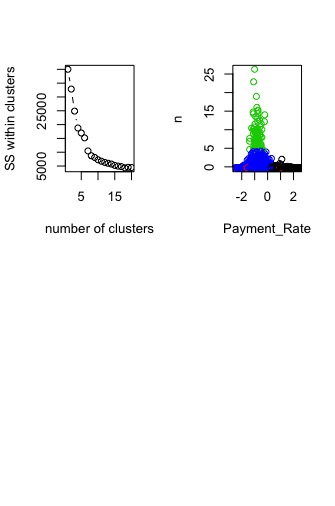
Take a brief look at the results, we can see that for cities which have much more medicare providers, the medicare cover rate is relatively low compared with some cities which have much less medicare providers. Why this situation happens? It is important for people to put focus on this difference and incorporate this insight to change this situation to improve people life quality.

**Data Exploration:**

** **

As we can see from these four pictures, all four variables distribution are right skewed. There are some outliers in these distributions. Then I can make a cor test . Based on the cor test, I select avg\_submitted, avg\_allowed, payment\_rate , n, as our clustering features.



Based on scree plot, I select the number of cluster as four. Based on the clustering plot, we find that for cities which have more medicare providers, the medicare cover rate( green cluster) is relatively low compared with cities which have less medicare providers ( black cluster). People should put focus on the green cluster to find why leads to this situation and fix the medicare cover rate into a moderate step to improve life quality of people.