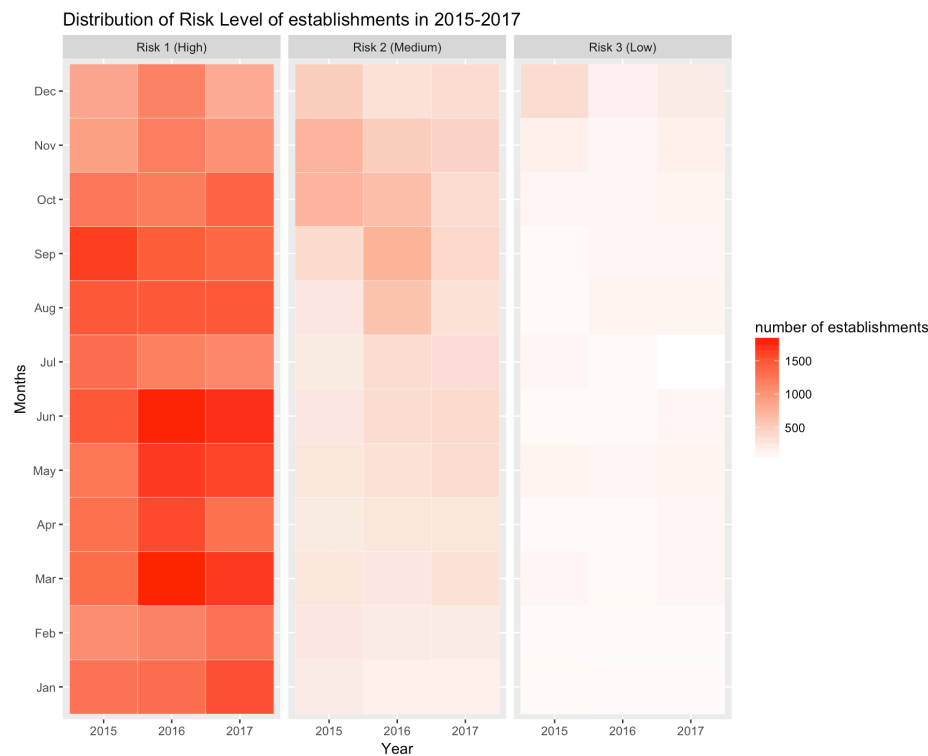


1. individual work

Dataset of our group is Chicago Food Inspection. I choose recent 3 years' data here.

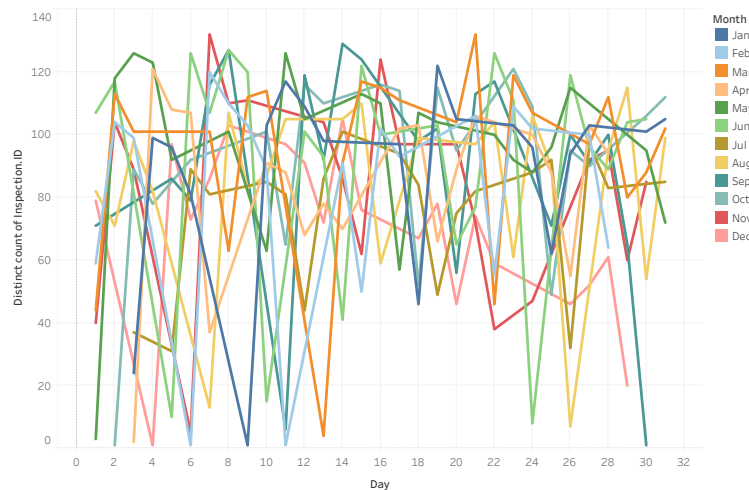
plot 1: Distribution of Risk Level of Facilities in 2015-2017



Through the graph, we can see that generally no matter in which year, the rank is Risk1 > Risk2 > Risk3, and from the colors we can see that the number of establishment in high risk take a really big part. As for year, the number of establishments of medium risk and low risk are yearly reduced which is a good news anyway.

plot 2: Food Inspection Frequency in 2017

Food Inspection Frequency in 2017

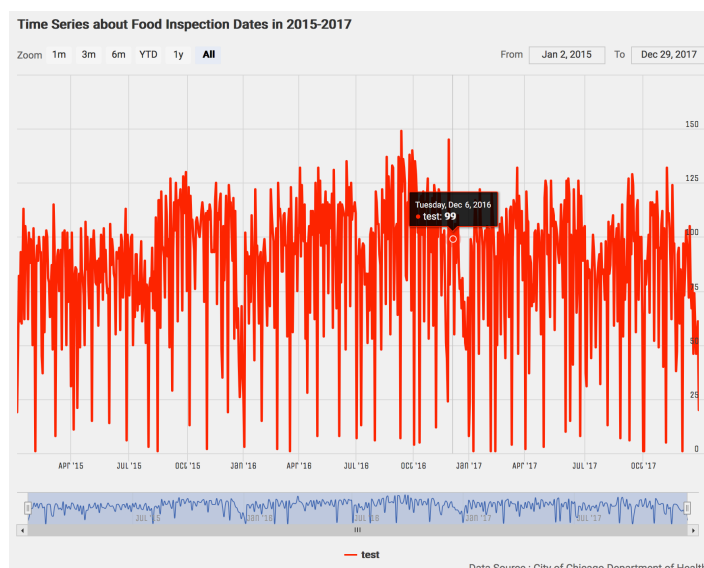


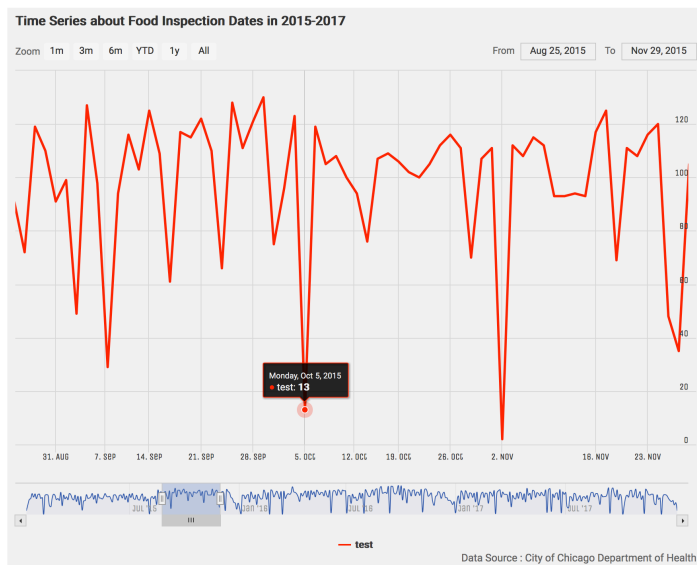
The trend of distinct count of Inspection.ID for Day. Color shows details about Month. The data is filtered on Year, which keeps 2017.

I use Tableau to map this overlay graph so even It looks clutter I can see each line individually and clearly.

Here we can get an interesting thing: inspectors seem would ‘take a break’ near every 5 days and inspection more happened in spring and winter and less in summer in general.

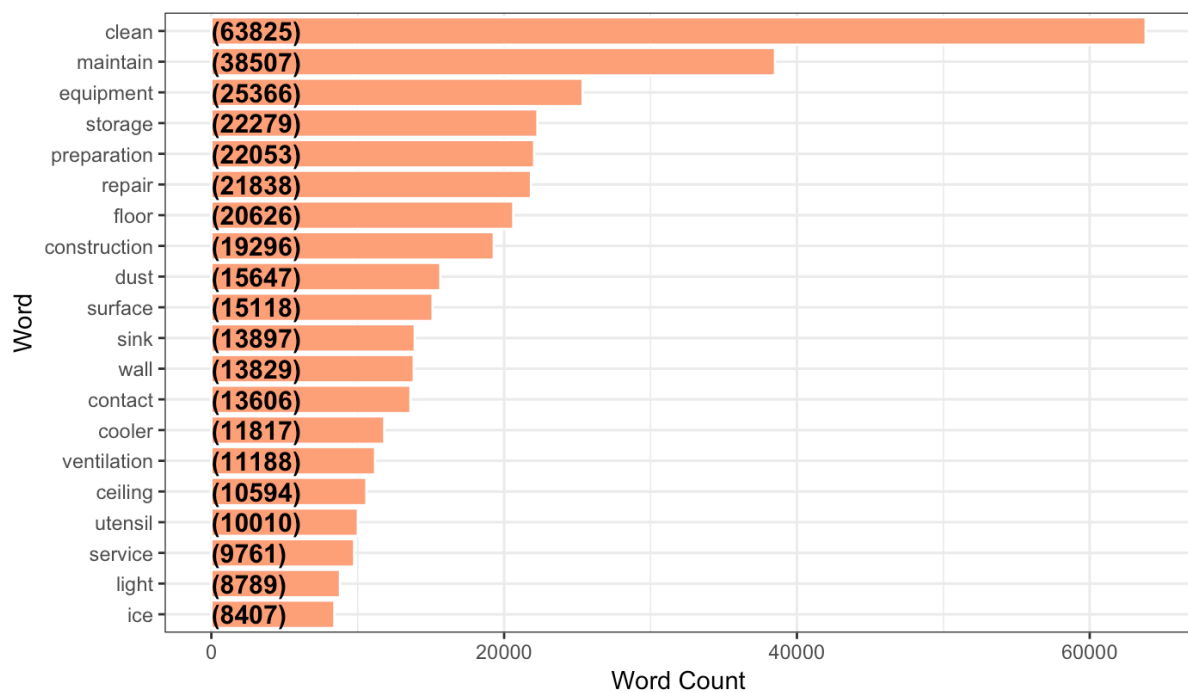
Besides, I create the time series interactive chart below which is easily to get 3-year information from without clutter.





#1.Top Twenty most Common Words

Top 20 most Common Words



#2. WordCloud of the Common Words



#Term Frequency of Words (TF)

