Problem Set #1

MACS 30200, Dr. Evans
Jingyuan Zhou

Data section

This dataset is obtained from the website of Chicago data portal. This information is derived from inspections of restaurants and other food establishments in Chicago from January 1, 2010 to the present (2017/4/15).

City of Chicago also used this data to help prioritize the city's food inspection staff to reduce the public's exposure to foodborne illness. It partnered with Allstate's Quantitative Research Analytics department to develop a predictive model to help prioritize the city's food inspection staff to reduce the public's exposure to foodborne illness. In this open-source project on github, their model evaluation calculates individualized risk scores for more than ten thousand Chicagoland food establishments.

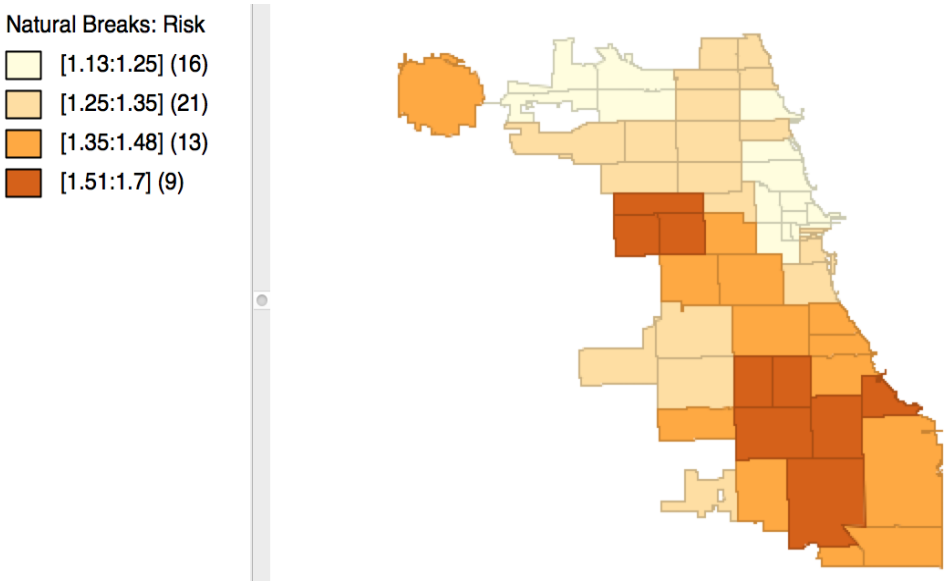
Data is collected by inspections that are performed by staff from the Chicago Department of Public Health's Food Protection Program using a standardized procedure. The results of the inspection are inputted into a database, then reviewed and approved by a State of Illinois Licensed Environmental Health Practitioner (LEHP).

Descriptive statistics of key variables are shown in the following tables:

	Facility Type	Zip	Inspection Type	Results	Violations
count	115239	115239	115239	115239	115239
unique	358	60	58	6	114231
top	Restaurant	60614	Canvass	Pass	18. NO EVIDENCE OF RODENT OR INSECT OUTER OPEN...
freq	78565	4191	59177	76382	31

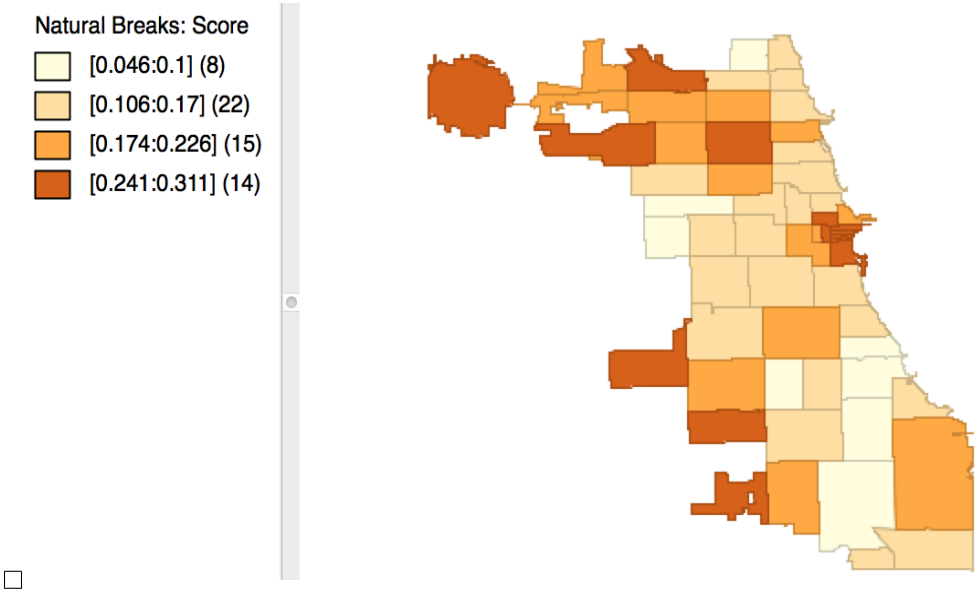
	Risk	Latitude	Longitude
count	115239.000000	115239.000000	115239.000000
mean	1.326391	41.878958	-87.676928
std	0.586120	0.082133	0.059652
min	1.000000	41.644670	-87.914428
25%	1.000000	41.820583	-87.708361
50%	1.000000	41.890944	-87.667042
75%	2.000000	41.939486	-87.634649
max	3.000000	42.021064	-87.525094

After grouping records by zipcode, the geographical distribution of the average risk level is shown below. Note that risk level of 1 is the highest level of risk and risk level of 3 is the lowest level of risk. Each establishment is categorized as to its risk of adversely affecting the public's health, and establishments with higher risk level are inspected more frequently. As we can observe from the map, northern part of Chicago generally have a higher risk, while southern part has lower risk.



Since the "Results" variable is categorical and is hard to visualize, I assigned 1 point to the records whose "Results" variable is "Pass"; -1 point to the records whose "Results" variable is "Fail"; 0.5 point to the records whose "Results" variable is "Pass w/ conditions". I ignored the records if their "Results" are "No Entry", "Not Ready" or "Out of Business" because they do not provide me much information about their inspection result. Thus, I've obtained one variable called "Score"

for each record. The geographical distribution of the average score of each zipcode area is shown below. We can observe that more zipcode areas in the north have relatively high scores compared to the south.



Apart from the geographical location of these establishments, different facility types could also have generally different risk level and score. After we group by facility types, the facility types that have lowest mean risk and highest mean score, in this case, with risk as 3 and score as 0.5, are:

```
['(convenience store)', 'BEVERAGE/SILVERWARE WAREHOUSE', 'CAR WASH',
'CHINESE HERBS', 'CONVENIENCE/GAS STATION', 'CONVNIENCE STORE',
'Convenience Store', 'DOLLAR STORE SELLING GROCERY', 'DRUG/GROCERY STORE',
'Dollar store', 'FOOD PANTRY', 'FROZEN DESSERT PUSHCARTS',
'GAS STATION STORE', 'GAS STATION/ GROCERY STORE',
'GAS STATION/GROCERY', 'GIFT/CARD SHOP WITH CANDY',
'GROCERY/LIQUOR STORE', 'GROCERY/SERVICE GAS STATION', 'GROCERY/TAVERN',
'Gas station', 'HERBAL MEDICINE', 'HERBAL STORE', 'HOOKA BAR',
'MOBILE DESSERTS VENDOR', 'MOBILE FOOD DESSERTS VENDOR',
'MOBILE FROZEN DESSERT VENDOR', 'Mobile Frozen Desserts Vendor',
'Mobile frozen dessert vendor', 'Nutrition Store',
'PREPACKAGE MEAL DISTRIBUTOR (1006 Retail)', 'PUSH CARTS',
'ROOM SERVICE', 'SERVICE BAR/THEATRE', 'SERVICE GAS STATION',
'SUMMER FEEDING', 'TAP room/tavern/liquor store', 'TAVERN/LIQUOR',
'TAVERN/RESTAURANT', 'TEA STORE', 'VFW HALL', 'night club', 'theater',
'video store', 'weight loss program']
```

Unsuprisingly, most of the facility types in this list are stores; however, mobile dessert vendors seem to also have better food inspection result.

The facility types that have highest mean risk and lowest mean score, in this case, with risk as 1 and score as -1, are

```
['Adult Family Care Center','BREWPUB',"CHILDERN'S SERVICES  FACILITY",
'DAYCARE 2 YRS TO 12 YRS', 'DINING HALL', 'SOUP KITCHEN', 'KIDS CAFE',
'MAIN KITCHEN', 'NOT FOR PROFIT','UNLICENSED FACILITY']
```

To our disappointment, some special care and non-profit facilities are in this short list.