Perspectives Research Problem Set 1

Julian McClellan

Due 4/19/17

Food Inspections Data

The data is drawn from inspections of restaurants and other food establishments in Chicago from January 1, 2010 to the present (April 13, 2017). The data encompasses 146,821 inspections.

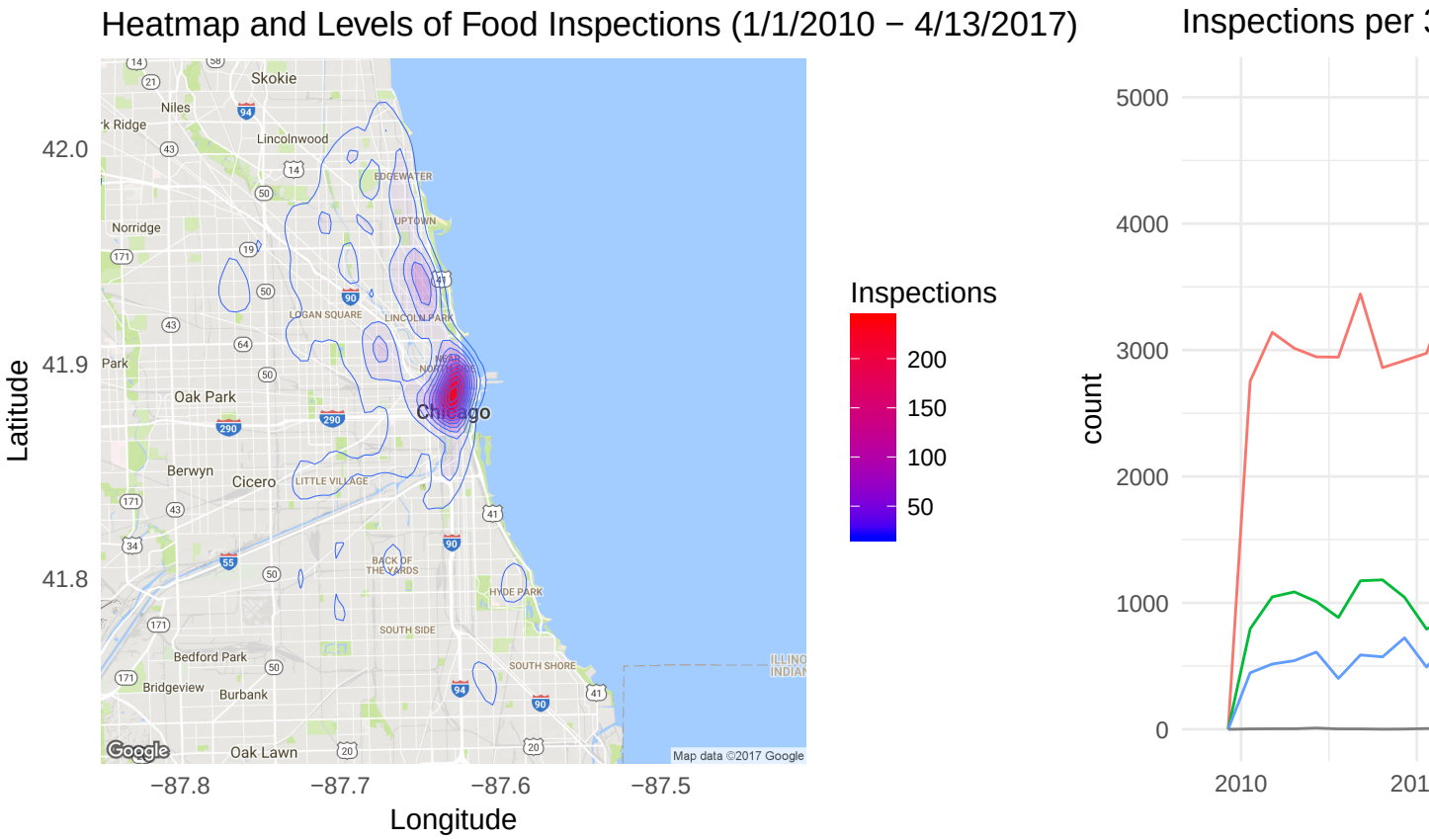
Inspections are performed by staff from the Chicago Department of Public Health's Food Protection Program using a standardized procedure. The results of the inspection are inputted into a database, then reviewed and approved by a State of Illinois Licensed Environmental Health Practitioner. The data is available on Chicago's open data portal. A detailed description of the variables is also available.

At a meeting of the American Public Health Association, (Schenk 2015) described an open source project in which:

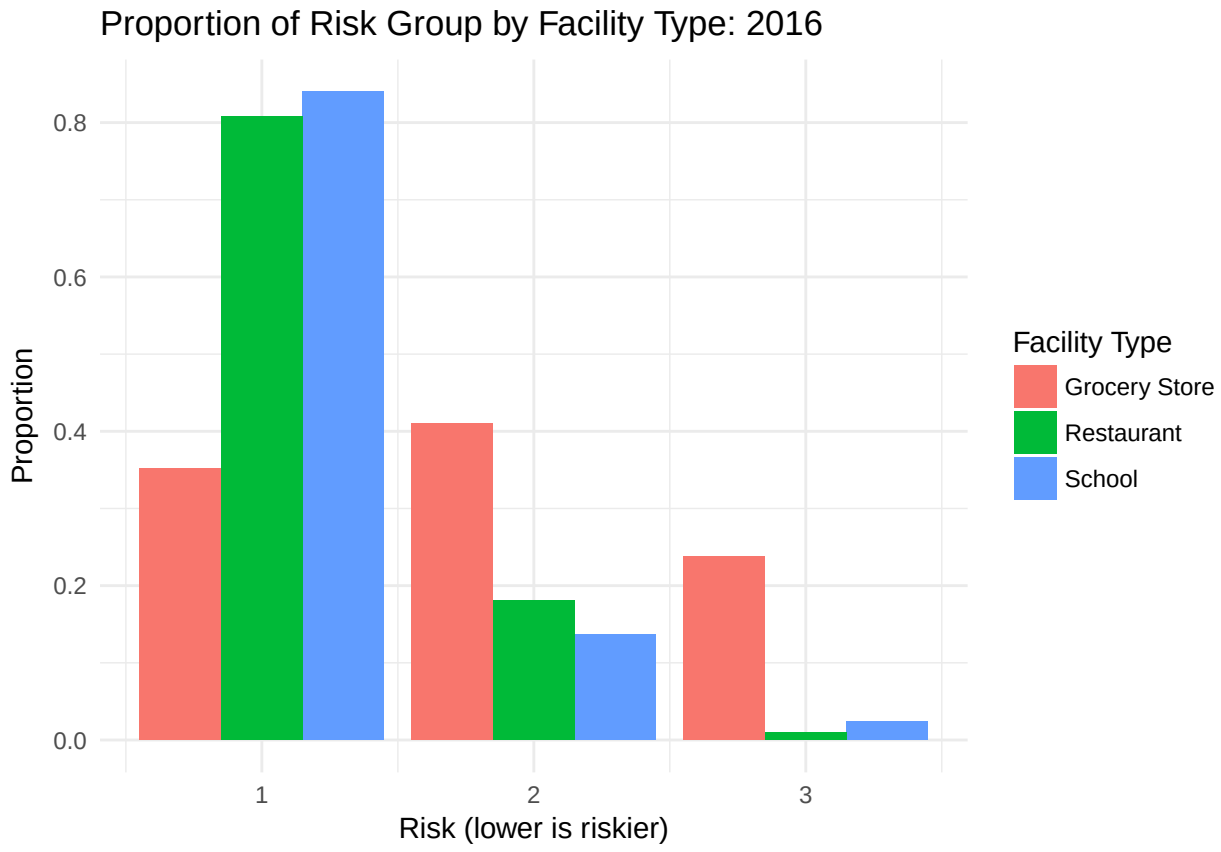
> [T]he City of Chicago's Department of Innovation and Technology (DoIT), in collaboration with an insurance company, and the CDPH, together developed advanced analytics to forecast food establishments that are most likely to have critical violations, which are most likely to contribute to food borne illness, so that they may be inspected first.

Making a *single* descriptive statistics table isn't exactly the most helpful thing given that the majority of the variables of categorical with differing numbers of categories, but here you go:

Categorical Variables	unique Values	# NA	Median inspections per unique value	Most Frequent Value
License #	32028	14	3	0
Legal Name	24134	0	4	SUBWAY
Public Name	23071	2696	4	SUBWAY
Facility Type	438	4540	4	Restaurant
Risk (lower = riskier)	4	79	22302	1
City	51	150	2	CHICAGO
Zip Code	100	98	1100	60614
Inspection Type	109	1	1	Canvass
Inspection Results	7	0	13407	Pass
Of course, this summary t	able can only te	ll us s	o much. What about where and when thes	e
inspections take place?				



However, one might also be interested in what facility types typically have higher risk, and in looking at specific years for the data. There are over 400 facility types, but let's compare restaurants, grocery stores, and schools in



2016:

Looking at this visualization, around 80% of restaurants and schools inspected in Chicago in 2016 were in the highest risk group (poor kids). Also, compared to the timeseries shown above, this visualization tells us that although a risk level of 1 (highest risk) is the most prevalent in the data, that this does not necessarily hold true within individual facility types. Also, along with the first summary table given, one might infer that the reason that risk level 1 food inspections dominate is due to the plurality (~67% of all observations) that restaurants hold in the data. One might expect then, that there are more nuances to tease out.

Schenk, Tom. 2015. "Food Inspection Forecasting to Optimize Inspections with Analytics." In *2015 Apha Annual Meeting & Expo (Oct. 31-Nov. 4, 2015)*. APHA.