Benjamin Rothschild
MACS 30200
Problem Set #1
Data Section Write Up

The dataset is derived from inspections of restaurants and other food establishments in Chicago from January 1, 2010 to the present and includes 146,955 observations.[1] Observations are collected from trained staff from the Chicago Department of Public Health's Food Protection Division[2] whose purpose is to promote public health in areas of food safety and sanitation. Inspections are done on food establishments such as restaurants, grocery stores, bakeries, hospitals, and schools among others. The department focuses their inspections on food handling practices, food temperature, personal hygiene, and pest control. For each inspection, an establishment can get a status of:

- **Pass**: establishment meets the minimum requirements of municipal codes and does not have any serious or critical violations
- **Pass with Conditions**: the establishment has Serious or Critical violations that are corrected during the inspection or the certified Food Service Sanitation Manager is not present as the time of the Inspection
- **Fail**: the establishment has Serious violations that cannot be corrected during the inspection. The business must correct the Serious violations promptly and pass a re-inspection to remain open. Note: the business can also have its license suspended until it passes re-inspection.

In addition, each establishment is given a Risk Code which categorizes the establishment as Risk 1 (high risk, with complex menus and food handling practices) to Risk 3 (low-risk, food is packaged or nonperishable). In general Risk 1 establishments are inspected twice per calendar year, Risk 2 once per year, and Risk 3 once every two years.

A description of the columns in the dataset

Column	Description	Sample Value
Inspection Number	Unique identifier for inspection	1286320
DBA Name	Business Name	CRUST EAT REAL
AKA Name	Business Alias	CRUST EAT REAL
License #	Unique value for license	87098
Facility Type	Type of facility	Restaurant, School, etc
Risk	Risk level of establishment	Risk 1, Risk 2, Risk 3
Address	Includes street, city, state, zip, latitude longitude	
Inspection Date	Date inspection was conducted	7/22/10

[1] Data can be accessed through the Chicago Data Portal (https://data.cityofchicago.org/Health-Human-Services/Food-Inspections/4ijn-s7e5/data) and is available in multiple formats (csv, json, xml) and through a Socrata REST API. Data was accessed April 19th, 2017

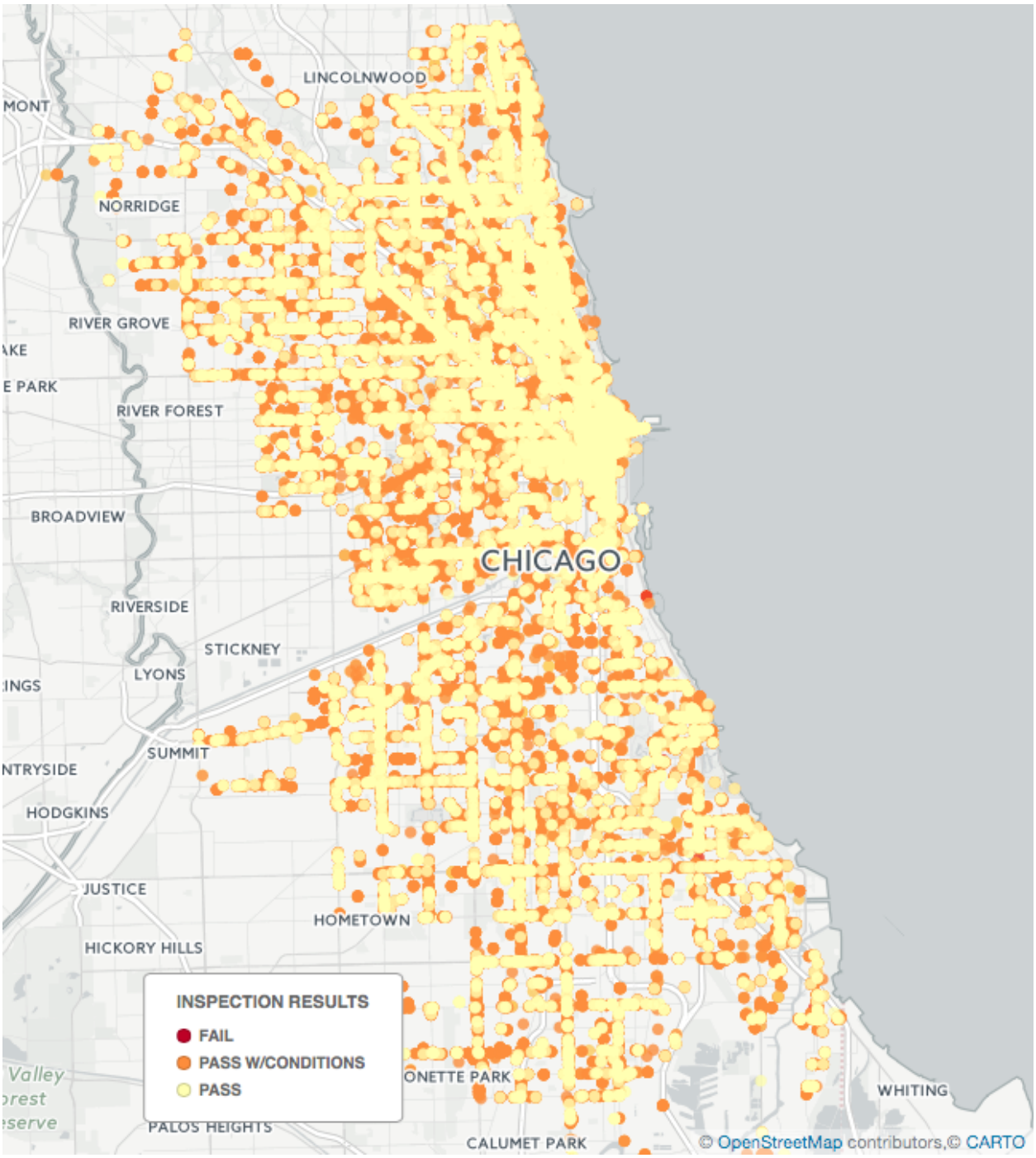
[2] http://webapps1.cityofchicago.org/healthinspection/General_Info.jsp

Inspection Type	Why the inspection was conducted	Canvas, Complaint, etc
Results	Results of inspection	Pass, Fail, Pass w/Conditions
Violations	List of violations	18. NO EVIDENCE OF RODENT OR INSECT OUTER OPENINGS PROTECTED/RODENT PROOFED, A WRITTEN LOG SHALL BE MAINTAINED AVAILABLE TO THE INSPECTORS - Comments: Violation #18 now corrected.

A summary of relevant dataset values are below.

Food Establishment Inspections January 1, 2010 - present	
Observations	146,955
Unique Establishments	32,053
Risk Level Results	
Risk 1	102,249
Risk 2	30,518
Risk 3	14,107
Missing	81
Inspection Results	
Fail	28,478
Out of Business	13,412
Pass	86,580
Pass w/Conditions	13,843
Other	30,518
Avg. Inspections Per Day	2,661

The coverage of the inspections are detailed in the map below.



Previous research using this data source tried to predict which food establishments would have the most critical violations to prioritize what establishments food inspectors should go towards first.[3] The purpose of this study was to more efficiently allocate food inspectors across the

[3] The methodology for the study is explained here https://chicago.github.io/food-inspections-evaluation/ and the code is open sourced here https://github.com/chicago/food-inspections-evaluation/

15,000 food establishments in Chicago. In this study, a dozen variables had substantial relationship with the likelihood of an establishment failing a food inspection such as if the establishment has previous failed inspections or its location and nearby sanitation complaints made through 311. In addition, research has been done using similar datasets from other cities. For example, researchers have successfully merged food establishment inspection databases with online reviews to predict the likelihood of a food violations[4] and online review sites like Yelp have included food inspection results to improve their product.[5]

[4] Kang, Jun Seok, et al. *Using Text Analysis to Target Government Inspections: Evidence from Restaurant Hygiene Inspections and Online Reviews*. No. 14-007. Harvard Business School, 2013.

[5] Booth, Darryl. "Yelp partners with health departments to improve food safety." *Journal of environmental health* 76.8 (2014): 52-56.