

An aerial night view of a city skyline, likely Chicago, with numerous lights from buildings and streets. The image is blurred, creating a bokeh effect with bright spots of light against a dark background.

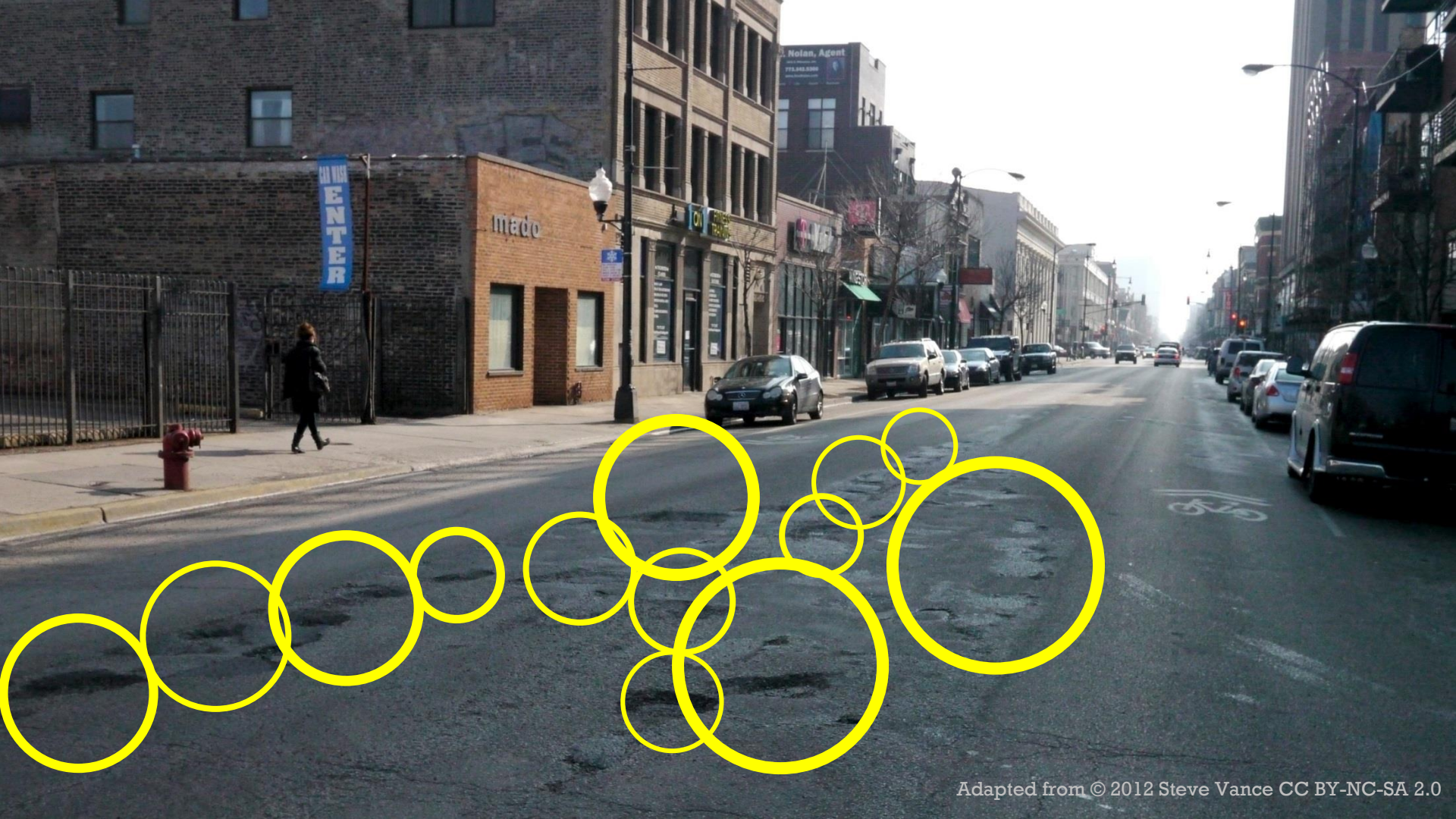
# CITY OF #BIGDATA

TOM SCHENK JR.  
CHIEF DATA OFFICER, CITY OF CHICAGO  
@CHICAGOCDO



**IN CHICAGO, WE BELIEVE THAT THE  
POWER OF TECHNOLOGY IS DRIVEN  
BY THE PEOPLE WHO USE AND  
BENEFIT FROM IT.**





**City of Chicago** Data Portal

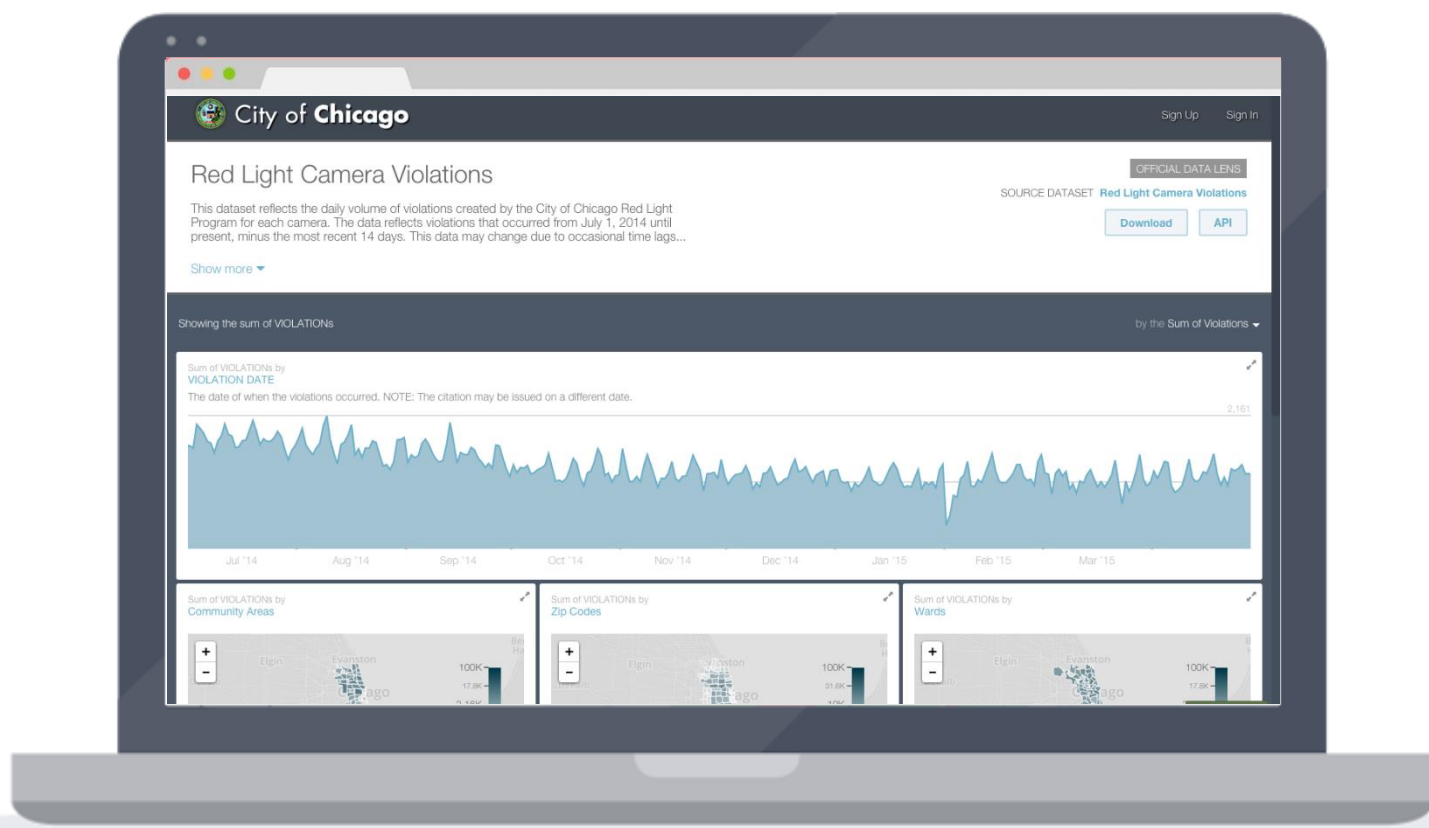
311 Service Requests - Pot Holes Reported

The Chicago Department of Transportation (CDOT) oversees the patching of potholes on over 4,000 miles of arterial and residential streets in Chicago.

	CREATION DATE	STATUS	COMPLETION DATE	SERVICE REQUEST NUMBER	TYPE OF SERVICE REQUEST	CURRENT ACTIVITY	MOST RECENT ACTION
1	11/14/2014	Open		14-01987391	Pothole in Street		
2	11/13/2014	Open		14-01985287	Pothole in Street		
3	11/13/2014	Open		14-01984596	Pothole in Street		
4	11/13/2014	Open - Dup		14-01985890	Pothole in Street		
5	11/13/2014	Open		14-01984723	Pothole in Street		
6	11/13/2014	Completed	11/13/2014	14-01985227	Pothole in Street	Final Outcome	Pothole Patched
7	11/13/2014	Open		14-01984083	Pothole in Street		
8	11/13/2014	Open		14-01983203	Pothole in Street		
9	11/13/2014	Open		14-01984802	Pothole in Street		
10	11/13/2014	Open		14-01983403	Pothole in Street		
11	11/13/2014	Open		14-01984109	Pothole in Street		
12	11/13/2014	Open		14-01983368	Pothole in Street		
13	11/13/2014	Open		14-01984335	Pothole in Street		
14	11/13/2014	Open - Dup		14-01983510	Pothole in Street		
15	11/13/2014	Open		14-01984368	Pothole in Street		
16	11/13/2014	Open		14-01984183	Pothole in Street		
17	11/13/2014	Open		14-01984745	Pothole in Street		
18	11/13/2014	Completed	11/13/2014	14-01985383	Pothole in Street	Final Outcome	Pothole Patched
19	11/13/2014	Open		14-01984782	Pothole in Street		
20	11/13/2014	Open		14-01980594	Pothole in Street		
21	11/13/2014	Open		14-01984378	Pothole in Street		
22	11/13/2014	Open		14-01985179	Pothole in Street		
23	11/13/2014	Open		14-01984870	Pothole in Street		
24	11/13/2014	Open		14-01983405	Pothole in Street		

Totals: 323428

Data on potholes are reported by residents and city staff through the 311 system, which is then reported on the City's #opendata portal—updated daily.

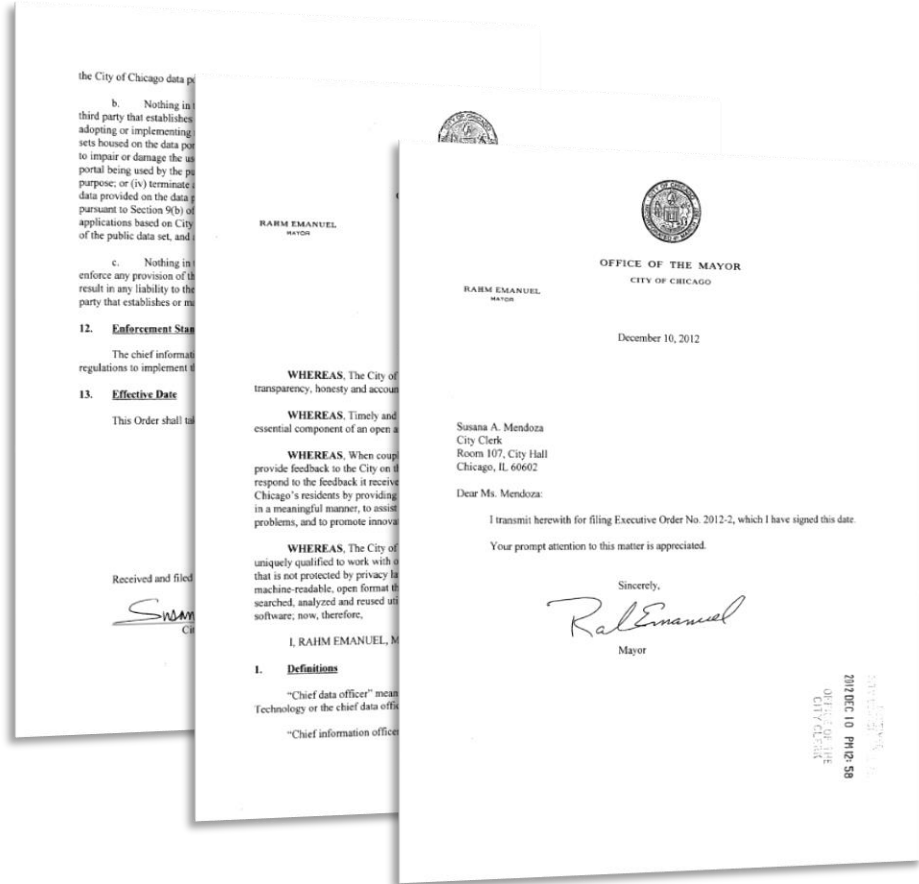


**Chicago has released more #opendata, including important items such as red light and speed camera violations, problem landlords, and public chauffeurs.**

[data.cityofchicago.org/view/caas-knxs](https://data.cityofchicago.org/view/caas-knxs)

## Executive Order 2012-2

In 2012, Chicago issued an executive order which formalized the open data portal, endowed powers to the Chief Data Officer, created an advisory committee to advise on the expansion of new datasets, and required an annual open data report.





LEVERAGE DATA AND NEW  
TECHNOLOGY TO MAKE  
GOVERNMENT MORE EFFICIENT,  
EFFECTIVE, AND OPEN



# INCREASE & IMPROVE CITY DATA

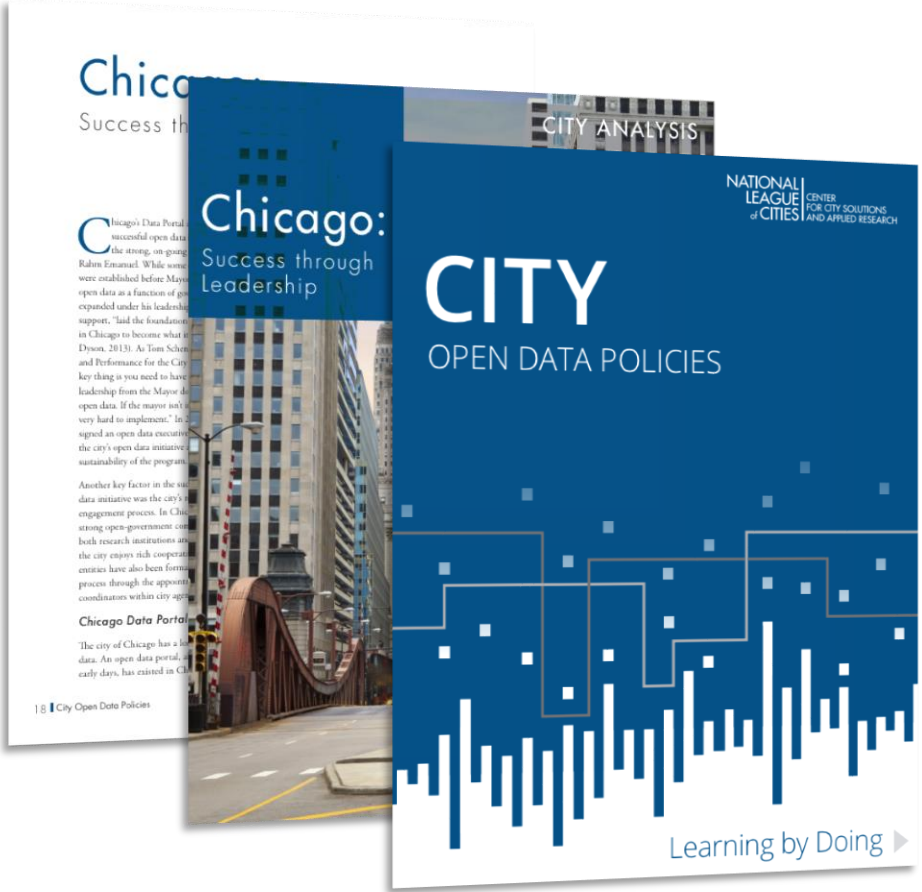
The City will continue to increase and improve the quality of City data available internally and externally, and facilitate methods for analyzing that data to help create a smarter and more efficient city.

**#OPENDATA PROVIDES A  
MEANS TO CREATE AN  
ECOSYSTEM AROUND DATA,  
WHICH INCLUDES MULTIPLE  
STAKEHOLDERS AND  
INITIATIVES THAT EXTEND  
BEYOND TRANSPARENCY.**

Open Gov Hack Night

happy **1** year!





# National League of Cities

NLC issued a report discussing the role of Chicago's leadership in developing a leading #opendata portal. The first chapter reviews Chicago's open data program and its benefits to the city, residents, and others.



© 2012  
National League of Cities

“Open data initiatives are an increasingly popular component of governance. At the national level, Chicago’s #opendata initiative has been held up as a model for cities that are seeking to start their own open data programs.”

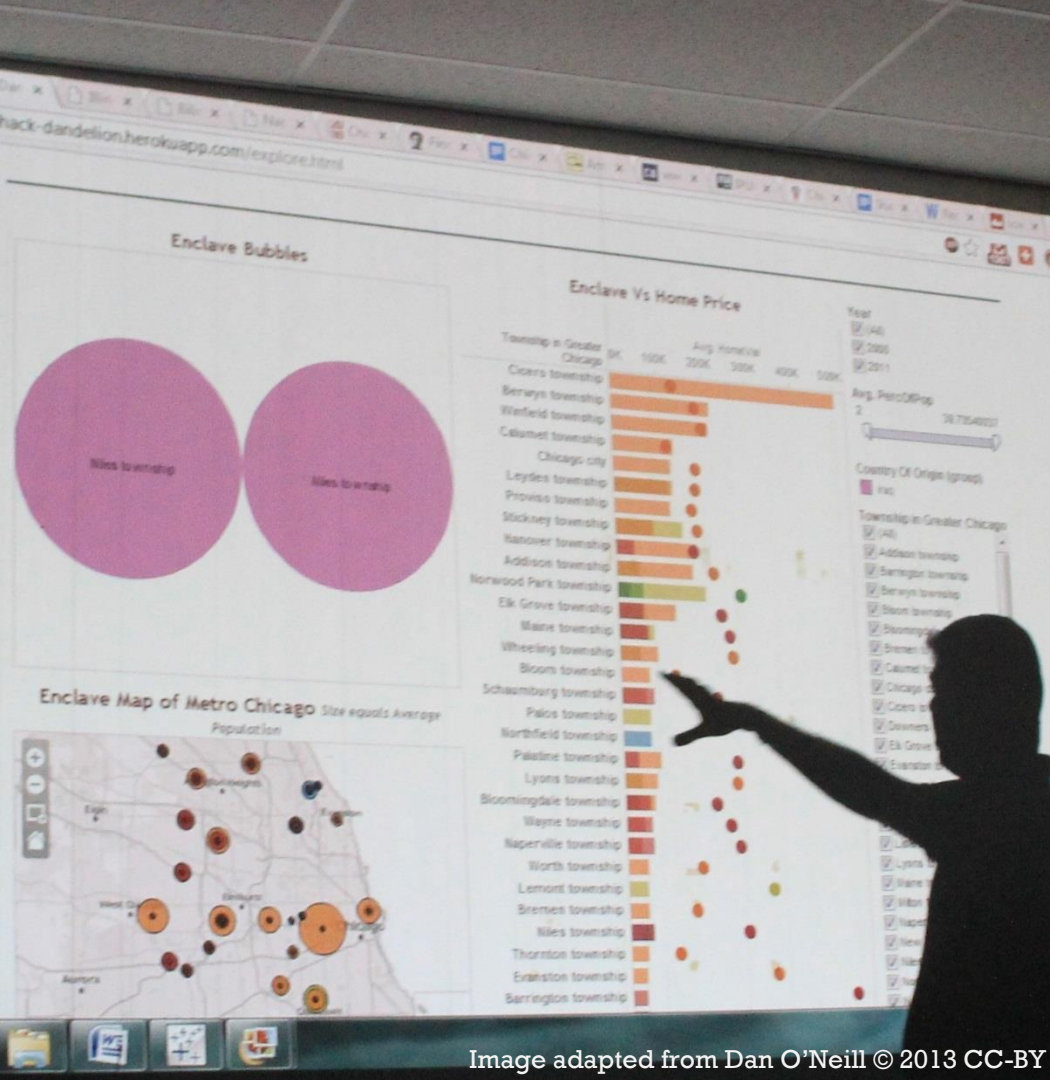
**- National League of Cities, p. 22**



© Tom Schenk Jr, 2016. CC-BY

## **Civic Tech Community**

Chicago has a large, vibrant, **productive**, civic community. This is led by Chicago residents interested in technology and society. Smart Chicago Collaborative and non-profits provide assistance and city officials regularly engage in meetups and other activities. This group has produced several helpful apps using #opendata.



# ChiHackNight

Meets every Tuesday

[chihacknight.org](http://chihacknight.org)

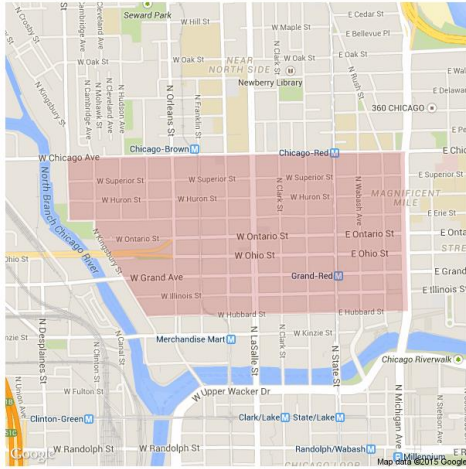
# City Data User Group

Meets first Weds of month

[meetup.com/ChicagoCityData](http://meetup.com/ChicagoCityData)

Ward 42, Sweep Area 2

ARE THE DATES OF YOUR NEXT STREET SWEEPING



**Want to be reminded of upcoming street sweepings?**

SweepAround.us has several options for being notified of street sweep in this area: we can send you an email, a text message or you may subscribe to a calendar feed. <sup>1</sup>

✉ Enter your email address, and we'll send alerts directly to your inbox.

e.g. you@example.com

Subscribe

📱 Receive text message alerts on your phone. To subscribe, text **W42A2** to **(773) 675-1985**. <sup>2</sup>

📅 Download or subscribe to this area's sweep calendar in iCal, Google Calendar or Outlook.

**2015 Sweep Schedule**

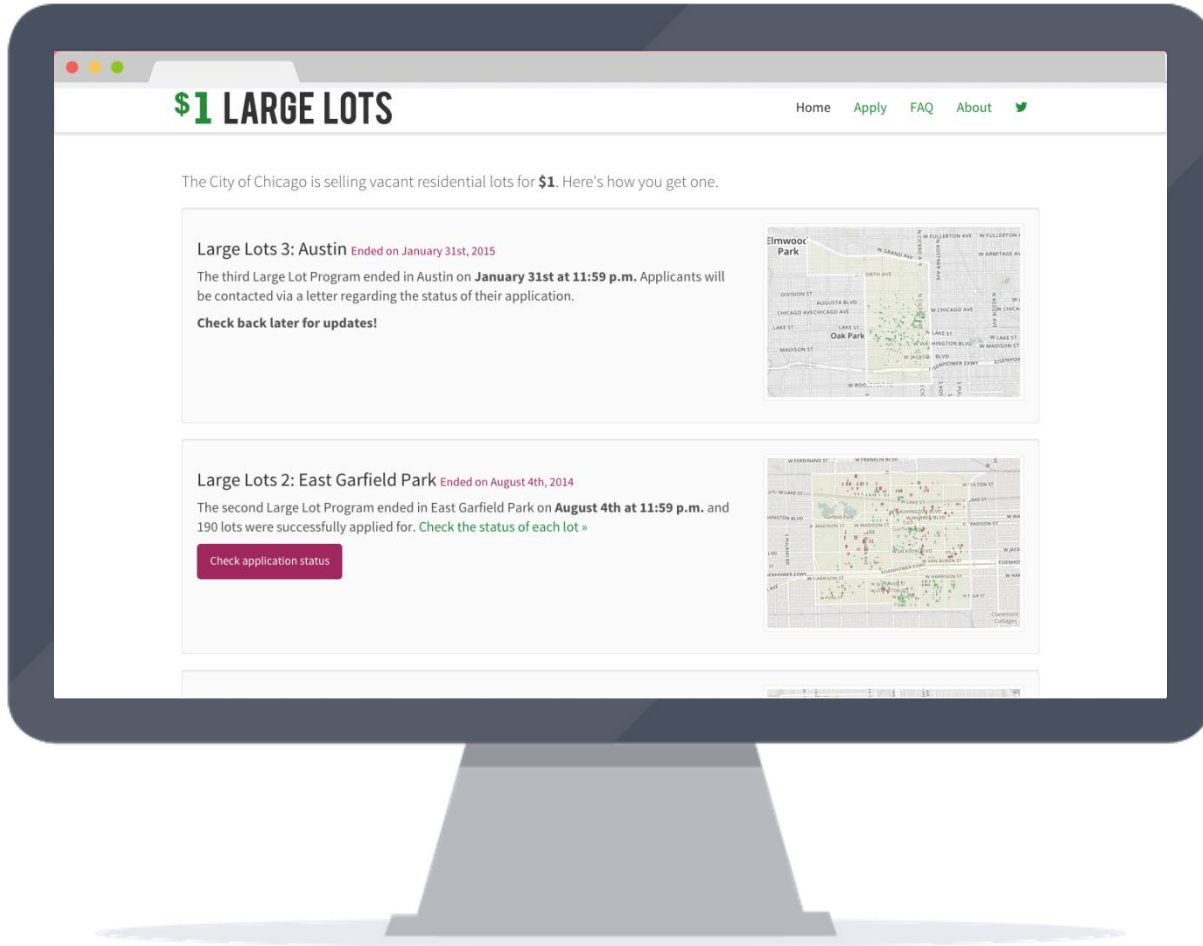
<sup>1</sup> Email and text message alerts will be sent one week, and one day, prior to the beginning of street sweeping in this area.

<sup>2</sup> Standard message and data rates may apply. To unsubscribe from alerts for this area, text STOP-W42A2.

**sweeparound.us**

Using #opendata, this service developed by the civic community alerts individuals to street sweeping activity by providing email, text, or calendar alerts.



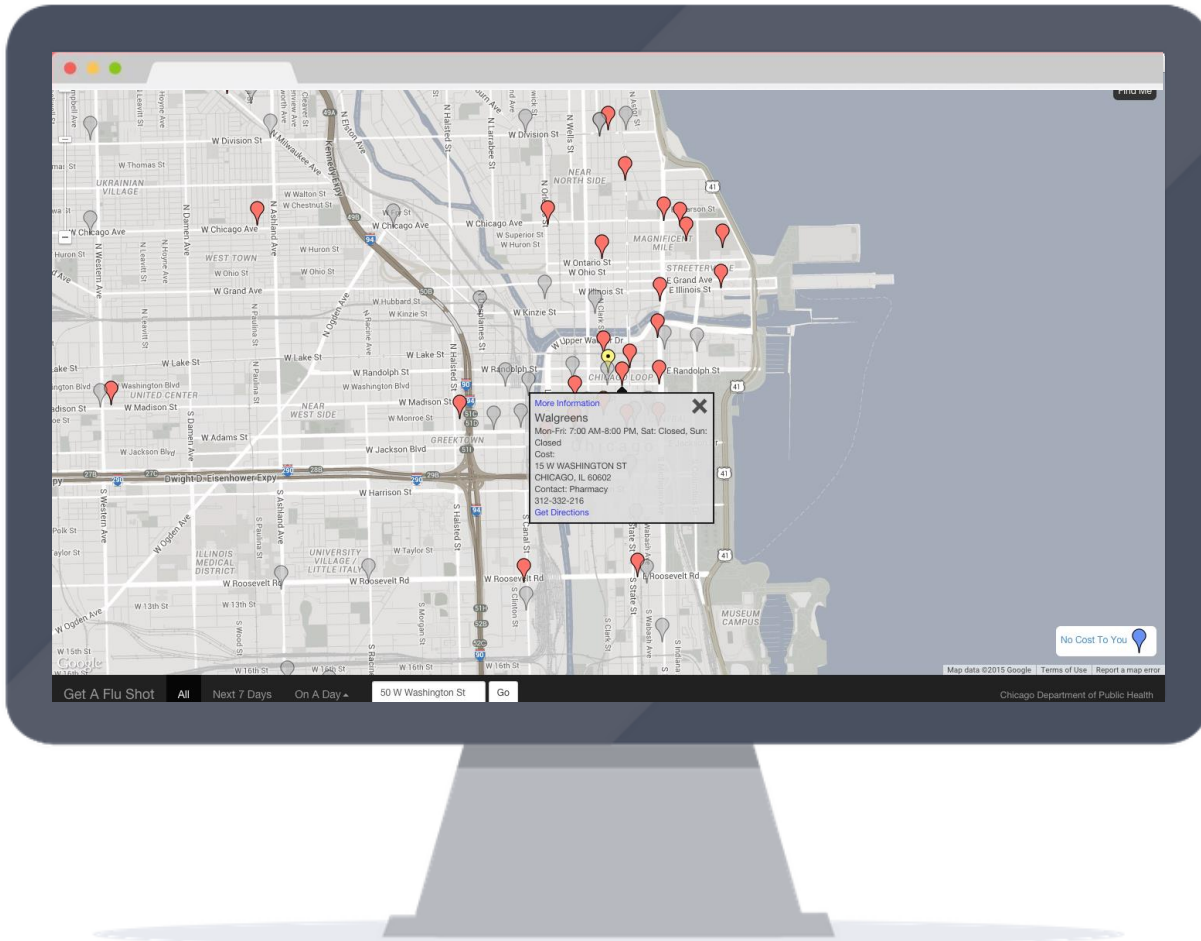


## largelots.org

The City of Chicago partnered with developers to create LargeLots, a website using #opendata to help residents apply to the City of Chicago \$1 lot program designed to encourage investment in struggling neighborhoods.

**chicagoflushots.org**

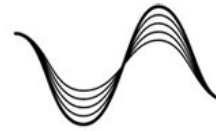
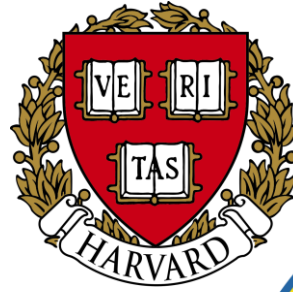
Chicago Flu Shots was developed to easily find flu-shot locations across Chicago. The code was created by a volunteer is #opensource so the site was adopted by Boston, Philadelphia, and San Francisco.



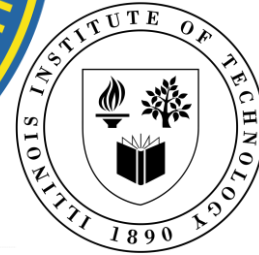
# Academia

---

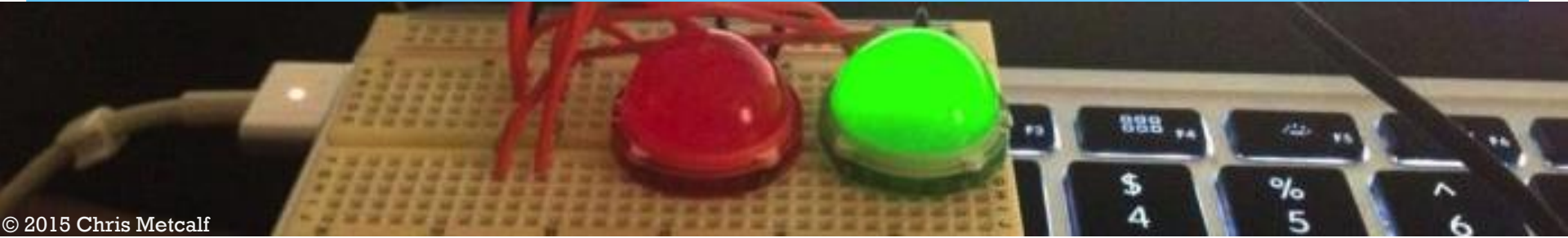
The City of Chicago has a number of high-quality research universities and groups willing to engage in projects with the city. We can leverage open data portal and data itself to create cooperative relationships.



Data Science  
for Social Good

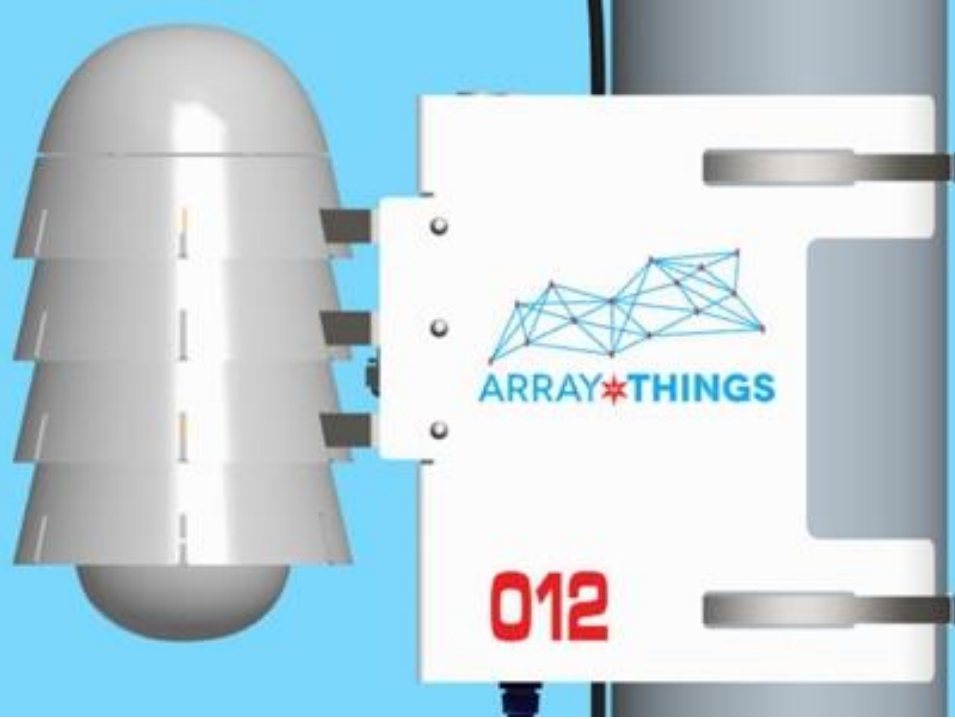


# OPEN DATA & INTERNET OF THINGS



© 2015 Chris Metcalf

Open data has also spread to physical devices. @chrismetcalf used traffic congestion data from the open data portal to generate an imp to provide a red or green light to denote heavy or light traffic congestion.



© 2015 University of Chicago

**Array of Things**  
[arrayofthings.github.io](http://arrayofthings.github.io)

University of Chicago has partnered with multiple institutions to build a mesh network of small sensors, dubbed the **Array of Things**, that will frequently post data for public consumption.

# Array of Things

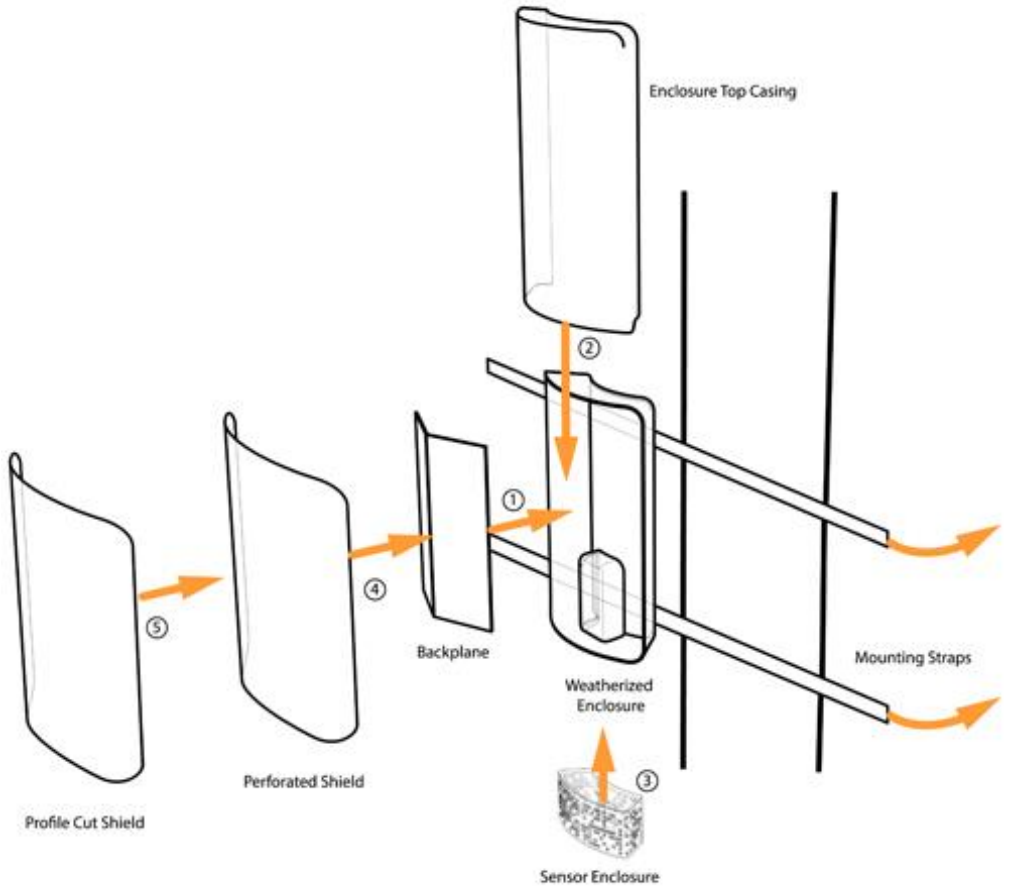
---



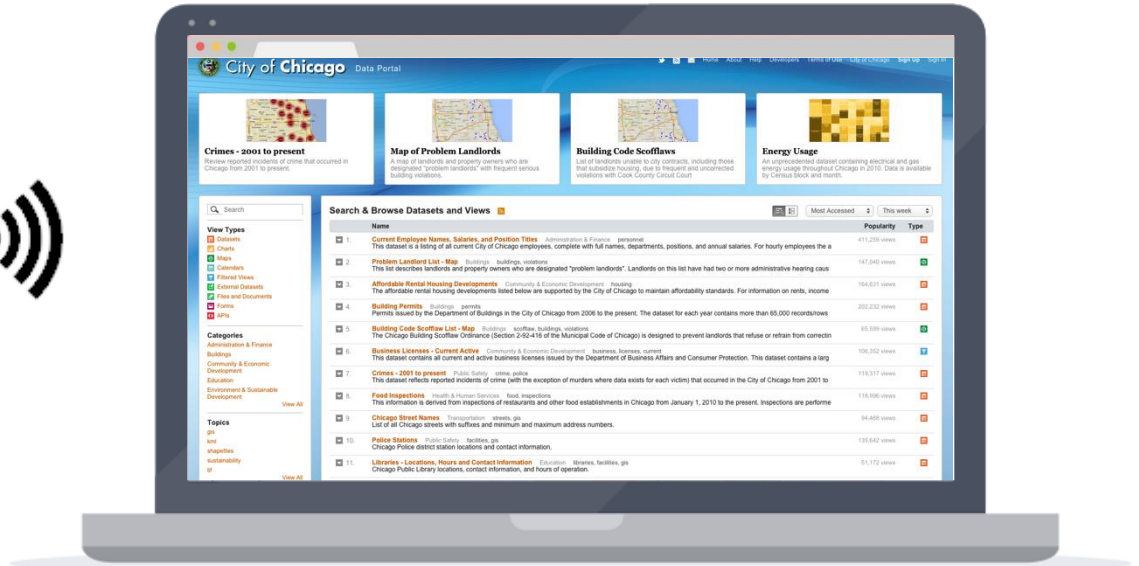
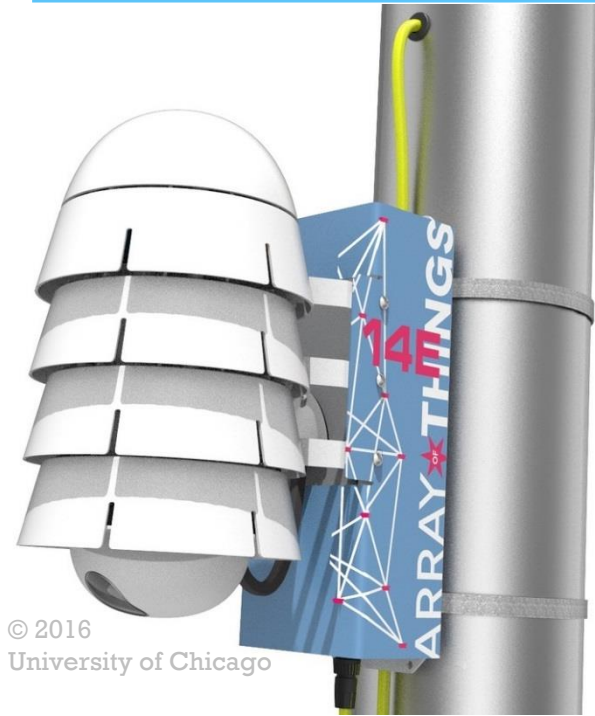
The Array of Things will provide hyper-local, temporal data on using a variety of sensors:

- **Sensors measuring sound and vibration**
- **Low-resolution infrared cameras** measuring sidewalk temperature
- **Climate and environmental data**, such as air-quality and temperature

# Array of Things



# OPEN INTERNET OF THINGS



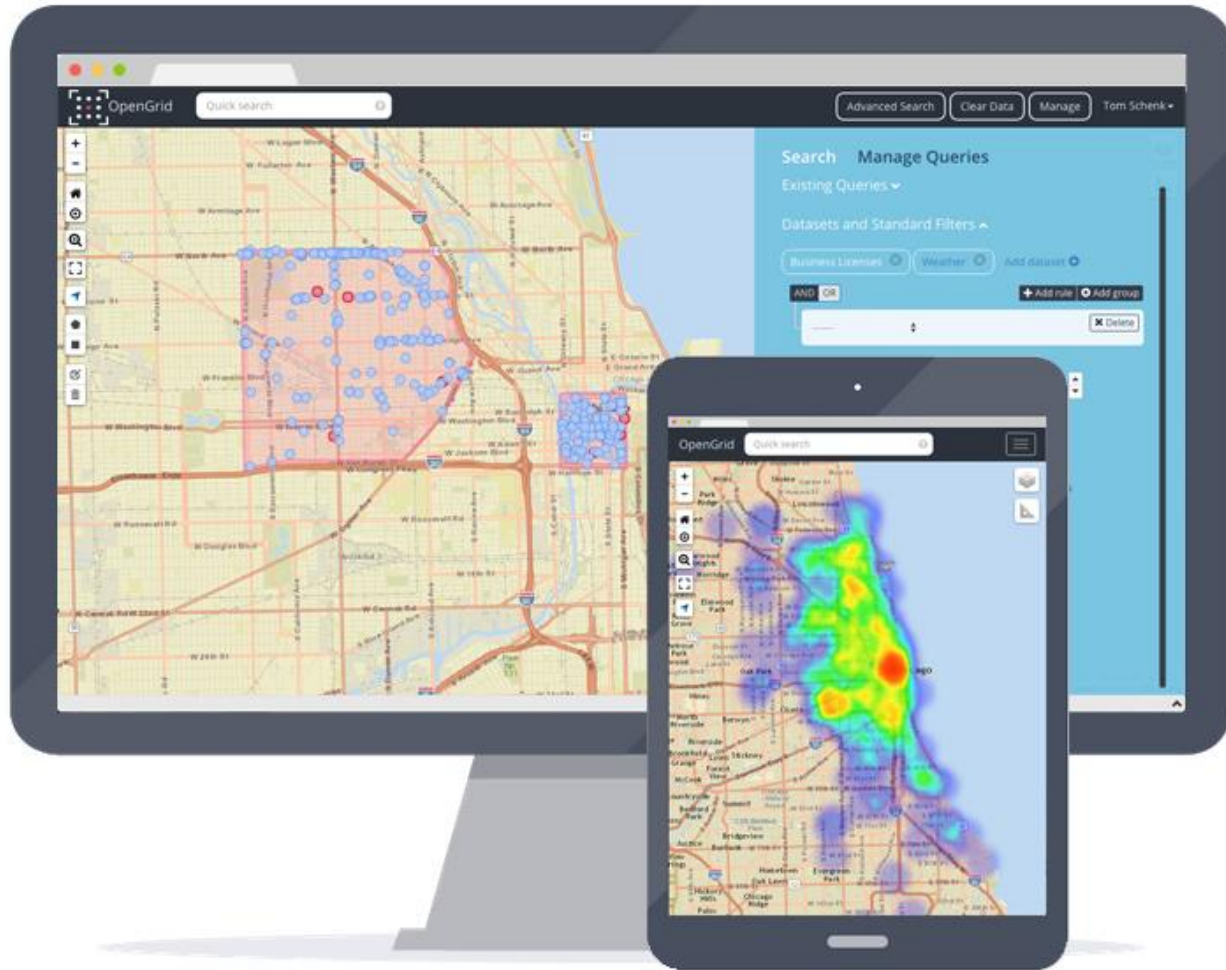
Array of Things

Chicago Open Data Portal









## OpenGrid.io

An #opensource platform which allows you to explore events such as 311 calls, crimes, permits, inspections, DIVVY trips in an interactive map. This software can be used by the public and an internal version drives situational awareness.

---

# ADVANCED ANALYTICS

---

## #Prediction

Using data and advanced research techniques to forecast and predict events in the city.

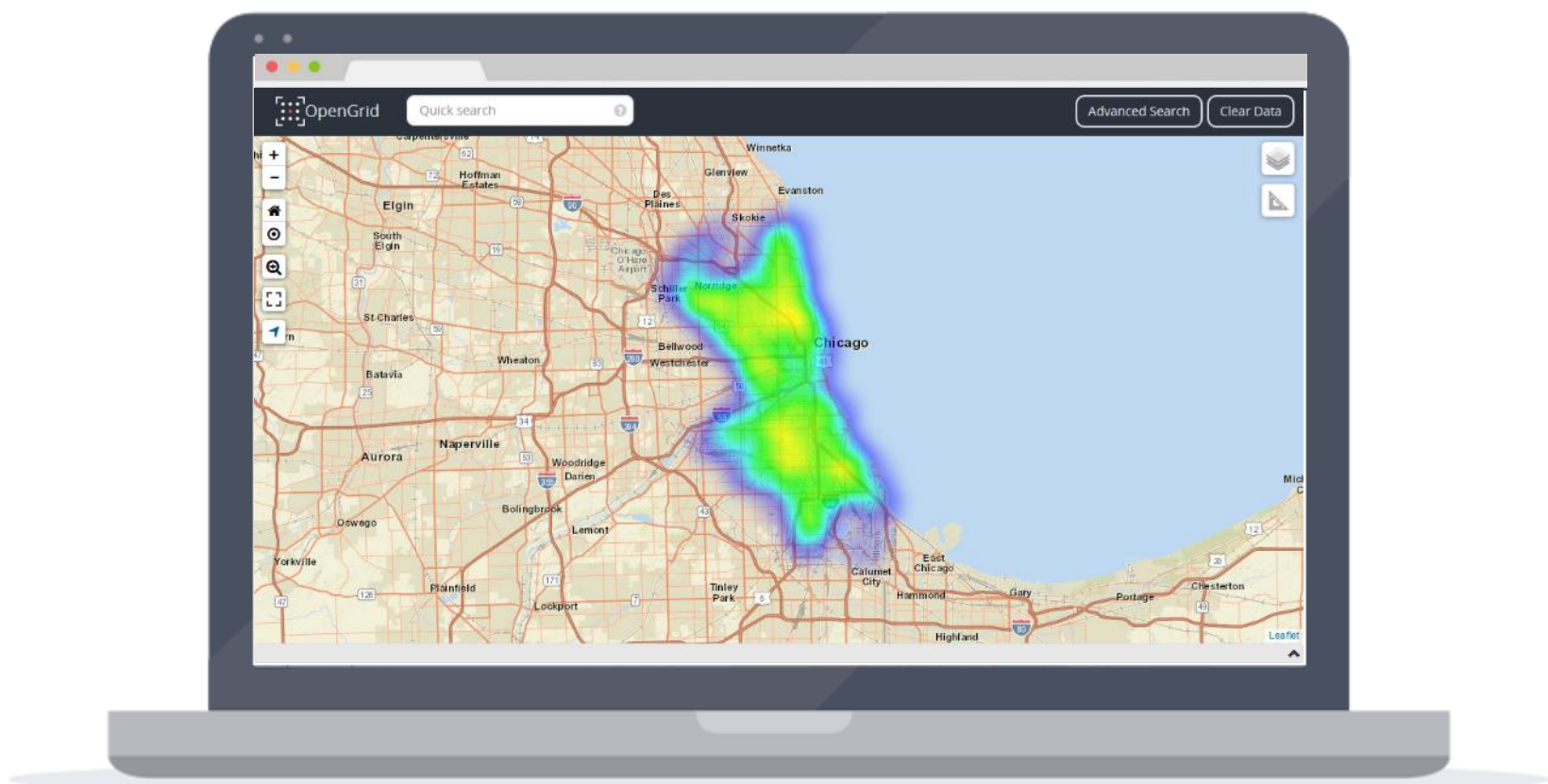
## #Evaluation

Evaluate the effectiveness of programs, including the effectiveness of advanced analytics.

## #Optimization

Optimizing the allocation of resources across the city to engage

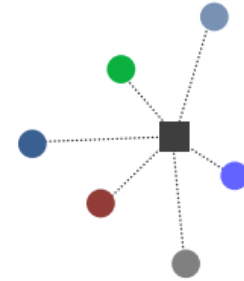




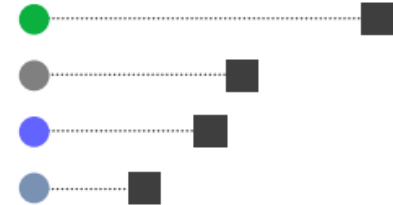
A map of rodent complaints across the city.

City of Chicago found 31 factors that predicted when and where rodent complaints are most likely in the next week. We used spatial-temporal relationships to create these #predictions, which started as an investigation of over 350 different factors.

Spatial Correlation



Temporal Correlation



Rodent Report for 06/24/2012 [Home](#) [Treatment Groups](#)

06/24/2012

District:24  
Score: 0.18123096000000005

Address	Block	Score	X	Y	District
2514_W_BIRCHWOOD_AVE	202001000	0.00581086	1158200	1949710	24
3029_W_GRANVILLE_AVE	207005005	0.00517836	1154910	1940950	24
6982_N_CLARK_ST	107001013	0.00428247	1163450	1946340	24
6327_N_ROCKWELL_ST	208001003	0.00409735	1157880	1942020	24
6306_N_HERMITAGE_AVE	303001007	0.00393256	1163590	1941760	24
2905_W_GRANVILLE_AVE	207002001	0.0300252	1155830	1940980	24
6226_N_FAIRFIELD_AVE	208008005	0.0125533	1156830	1941340	24
2923_W_ROSEMONT_AVE	207006006	0.012346	1155630	1941620	24
6317_N_MOZART_ST	207001001	0.0114892	1158230	1941670	24
6052_N_FAIRFIELD_AVE	208007005	0.0104587	1156860	1940290	24
6450_N_LEAVITT_ST	205001010	0.00930737	1160420	1943000	24
6130_N_TALMAN_AVE	208002003	0.00258584	1157510	1940740	24
2711 W FSTERS AVF	203004003	0.00258159	1156980	1946050	24

The #predictions generate a list of likely locations and published to an internal site used to route preventative baiting crews to bait likely locations.

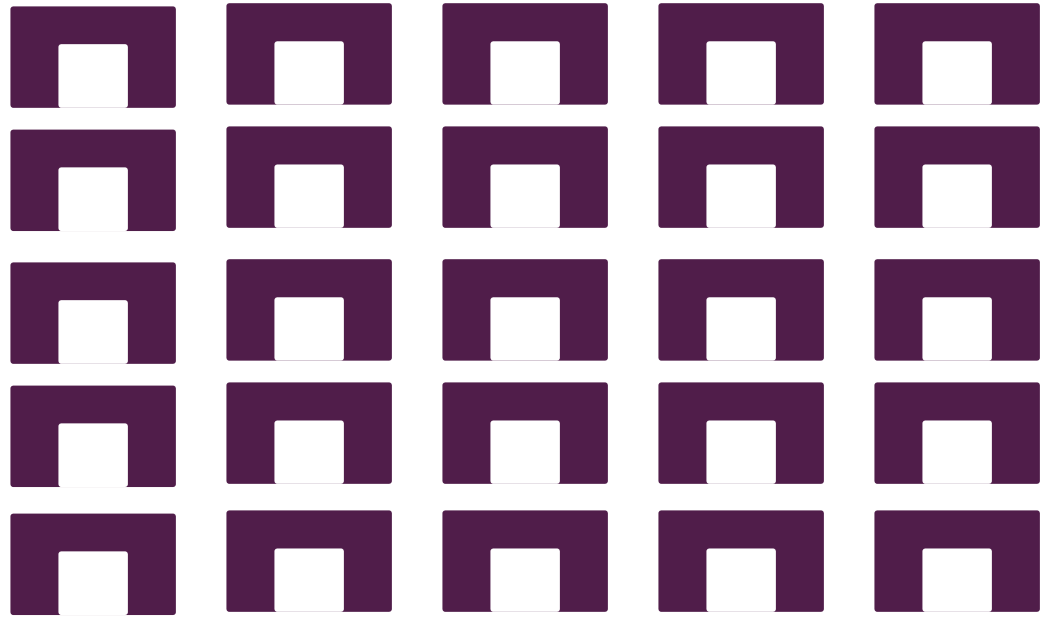
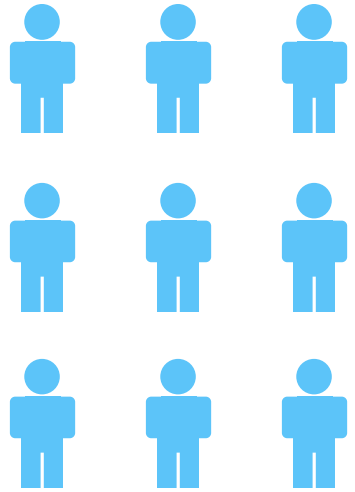




---

# OPTIMIZING FOOD INSPECTIONS

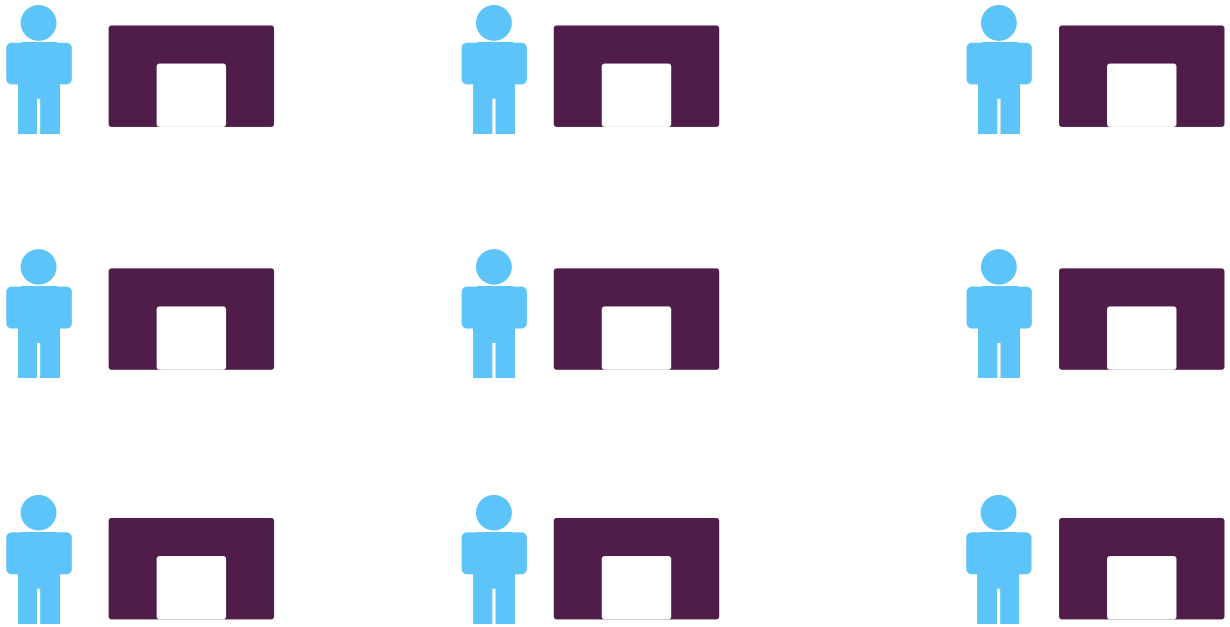
---



---

# OPTIMIZING FOOD INSPECTIONS

---





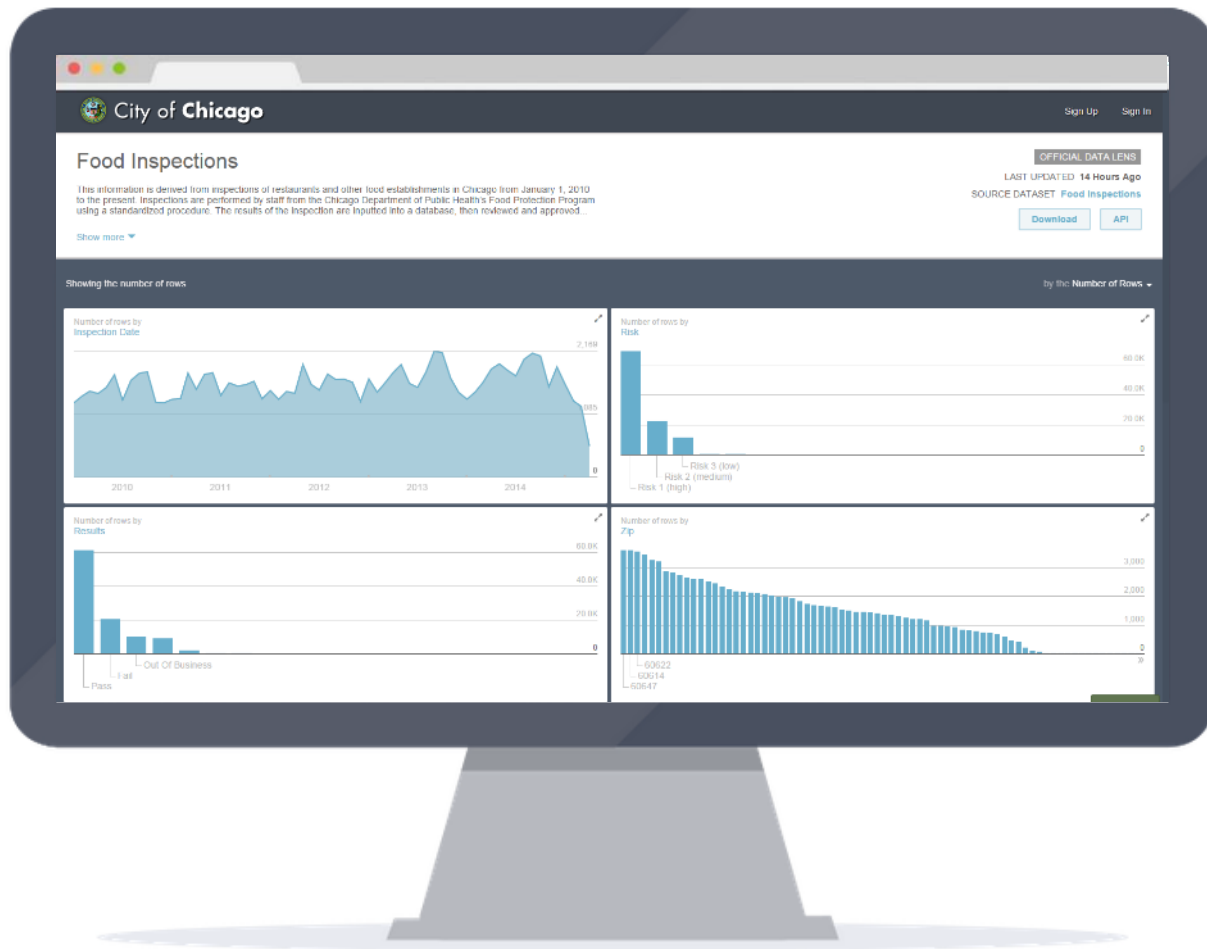
23%

11%

7%

15%

14%



## USING OPEN DATA

Chicago leveraged the #opendata portal to share data with external researchers, leveraging the city's premiere method of sharing data and saving time on data-sharing agreements to create #predictions.

The model predicts the likelihood of a food establishment having a critical violation, a violation most likely to lead to food borne illnesses. Over a dozen #opendata sources were used to help define the model. Ultimately, ten different variables proved to create #predictions of critical violations.

## **Significant Predictors:**

Restaurants with previous critical violations

Three-day average high temperature

CDPH risk level

Location of restaurant

Nearby garbage and sanitation complaints

Type of facility

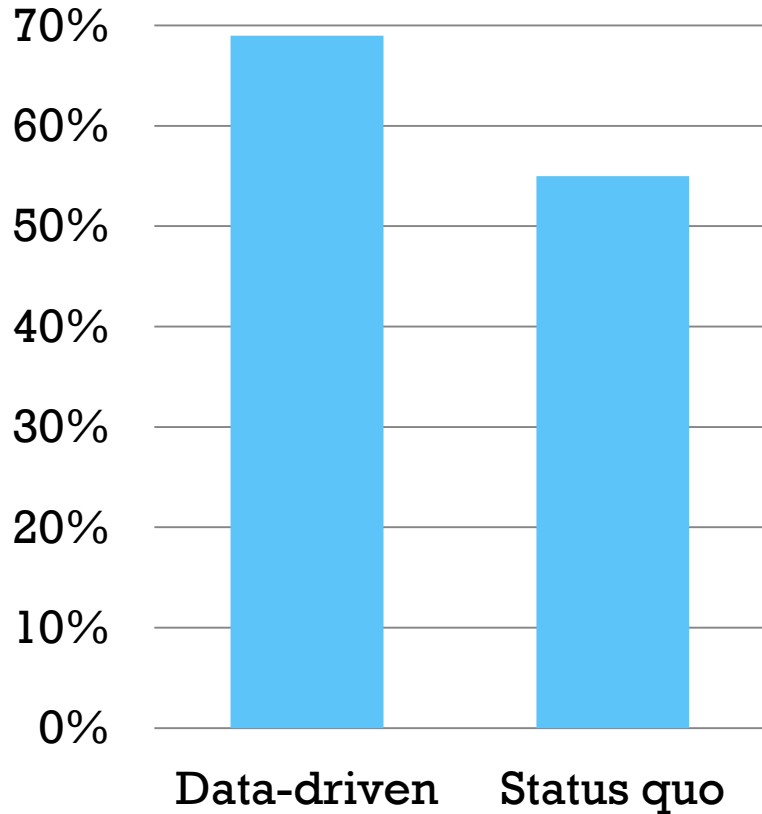
Nearby burglaries

Whether the establishment has a tobacco or has an incidental alcohol consumption license.

Length of time since last inspection.

Length of time the restaurant has been inspecting.

## Critical violations



The #predictions revealed an opportunity to find deliver results faster. Within the first half of work, 69% of critical violations would have been found by inspectors using a data-driven approach. During the same period, only 55% of violations were found using the status quo method.

---

## IMPROVEMENT

# 7 days

**The food inspection model is able to deliver results faster.**

After comparing a data-driven approach versus the status quo, the rate of finding violations was **accelerated** by an average of 7.4 days in the 60 day pilot. That means the #predictions led to more violations would be found sooner by inspectors.

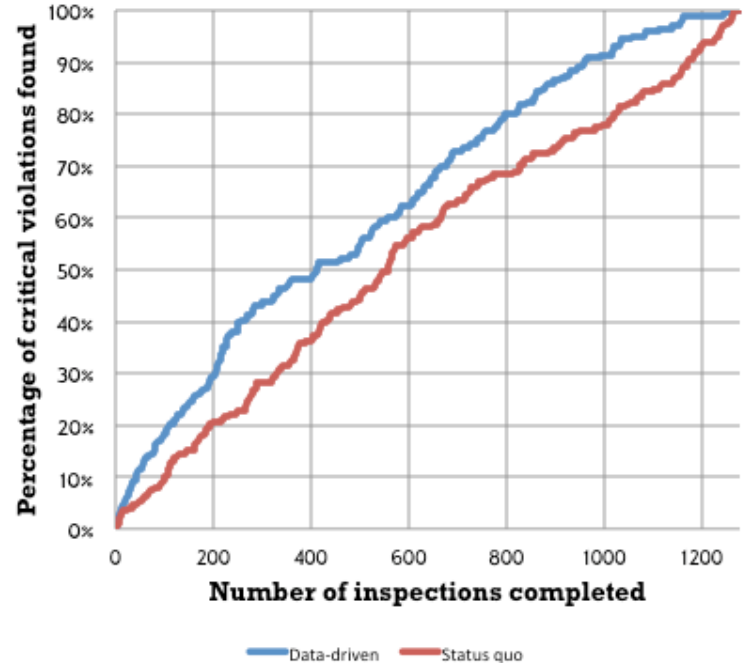
---

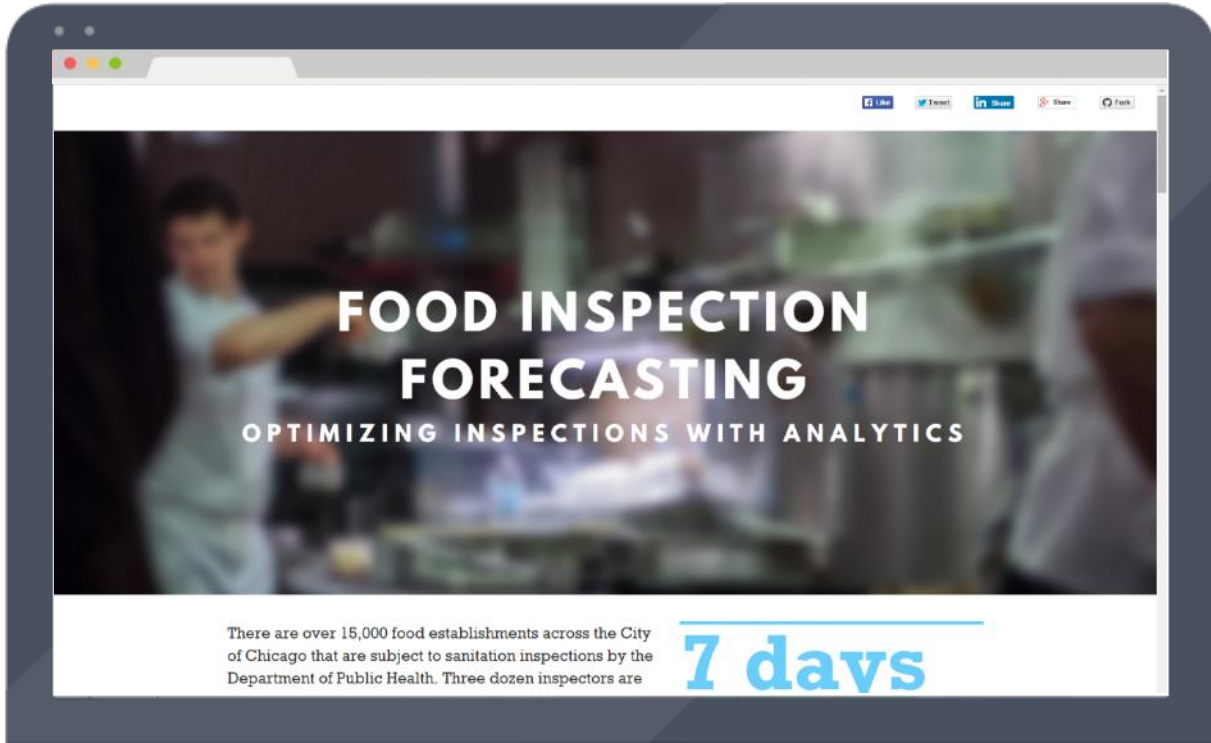


# OPTIMIZING FOOD INSPECTIONS

Discovering critical violations sooner rather than later reduces the risk of patrons becoming ill, which helps reduce medical expenses, lost time at work, and even a limited number of fatalities.

## Impact





There are over 15,000 food establishments across the City of Chicago that are subject to sanitation inspections by the Department of Public Health. Three dozen inspectors are

**7 days**

<http://chicago.github.io/food-inspections-evaluation/>

Chicago Food Inspections

Restaurants Inspectors

Show 25 entries Search:

Click to Download Filtered Data

Download

Filter by Zip Code

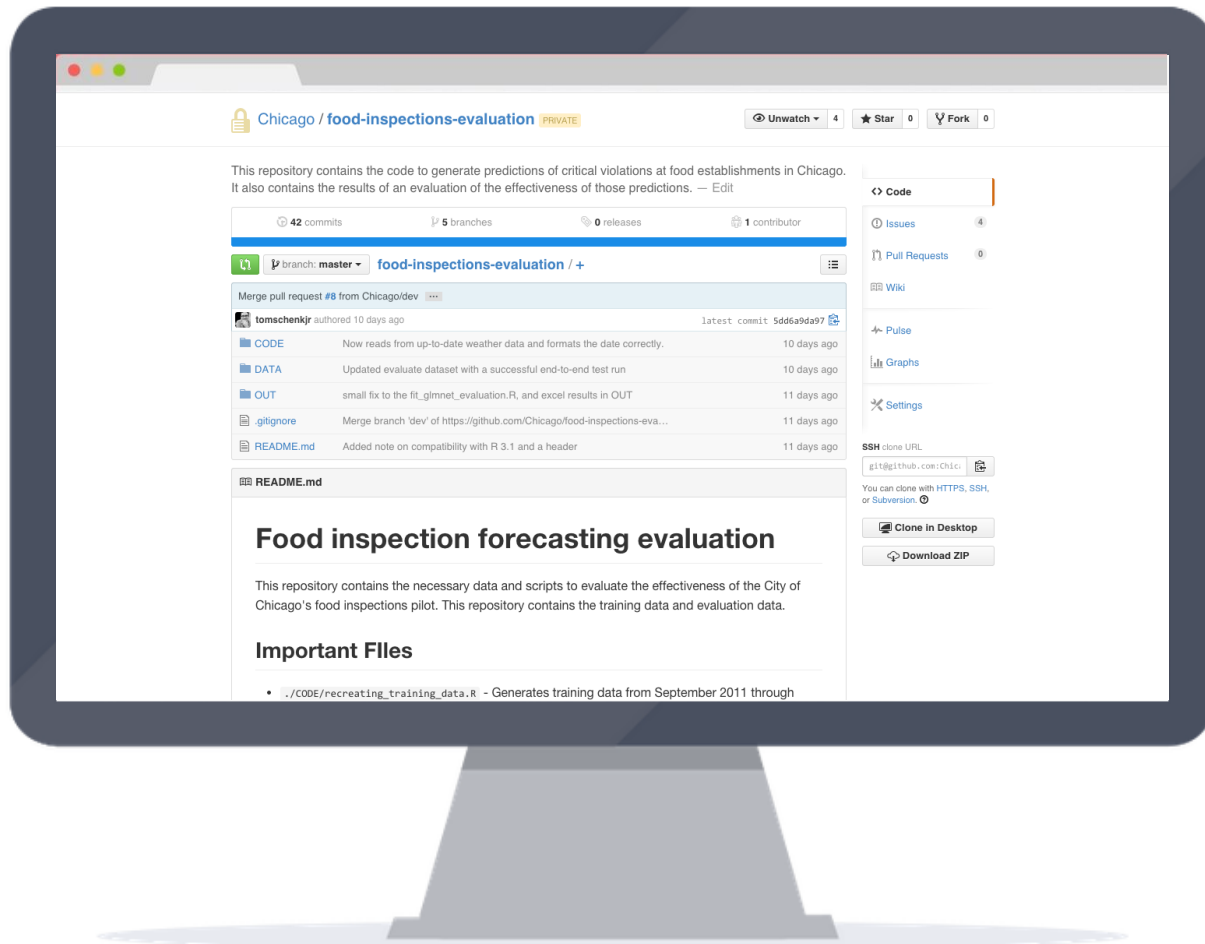
all

Filter by Inspector

all

License	Doing Biz As	AKA Name	Facility Type	Risk	Address	City	Zip	Prediction	License Code
2	COSI	COSI	Restaurant	Risk 1 (High)	230 W MONROE ST	CHICAGO	60606	0.09423278	1005
9	XANDO COFFEE & BAR / COSI SANDWICH BAR	XANDO COFFEE & BAR / COSI SANDWICH BAR	Restaurant	Risk 1 (High)	116 S MICHIGAN AVE	CHICAGO	60603	0.16524516	1005
40	COSI	COSI	Restaurant	Risk 1 (High)	233 N MICHIGAN AVE	CHICAGO	60601	0.09158854	1005
92	XANDO COFFEE & BAR / COSI SANDWICH BAR	XANDO COFFEE & BAR / COSI SANDWICH BAR	Restaurant	Risk 1 (High)	230 W WASHINGTON ST	CHICAGO	60609	0.09011859	1005
99	XANDO COFFEE & BAR / COSI SANDWICH BAR	COSI	Restaurant	Risk 1 (High)	203 N LA SALLE ST	CHICAGO	60601	0.09994210	1005
115	JOHN SCHALLER	JOHN SCHALLER	Restaurant	Risk 1 (High)	3714 S HALSTED ST	CHICAGO	60609	0.09250459	1005
149	FOX'S BEVERLY PUB	FOX'S BEVERLY PUB	Restaurant	Risk 1 (High)	9586 S WESTERN AVE	CHICAGO	60643	0.12944078	1005
158	BURWOOD TAP	BURWOOD TAP	Restaurant	Risk 2 (Medium)	724 W WRIGHTWOOD AVE	CHICAGO	60614	0.13499741	1005
164	POTASH BROS SUPERMARKET	POTASH BROS SUPERMARKET	Grocery Store	Risk 1 (High)	1525 N CLARK ST	CHICAGO	60610	0.16105854	1005
207	RIVLR SHANNON	RIVLR SHANNON	Restaurant	Risk 2 (Medium)	425 W ARMITAGE AVE	CHICAGO	60614	0.08928112	1005

The data science team has built a website which lets CDPH prioritize inspections based on projected risk.



## #OPENSOURCE

The analytical model will be released as an open source project on GitHub, allowing other cities to study or even adopt the model in their respective cities. No other city has released their analytic models before this release.

<http://github.com/Chicago/food-inspections-evaluation>

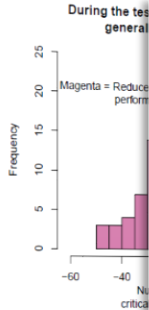
# Technical Documentation

The project was released using an academic-quality technical paper instructing others on the the variables and statistical methodology used in the project. In addition to source code, the paper will help researchers adopt this approach.

Education, weather is including  
one a single temperature with  
we're only moving the internal  
predicted value of a food rate  
Additional data can also be  
Flavor A/P or V, could be  
systematic model, outside  
the model presented in this  
report. The instructions to

## References

Friedman, Jerome, Teresa Elia,  
Mokhe vs. Coordinate Dances  
80/7/  
Harris, James K., Royal Mass  
Department of Social Med  
Mortality, Weekly Report 63  
Simon, Noah, James Frank  
Cox's Proportional Hazards  
<http://www.johndoff.org/v2/>  
Yonahin, W. N., and H. D.  
<http://www.state-on-a-id.jp>



We conducted a t-test to measure vs  
zero. Namely, the null hypothesis is  
The test ( $n = 25, 20, df = 257$ ) result  
likely to be significant.

Below, Gini curves show the relative  
first day, the data-driven model re  
found in the first week between Sep  
for business-as-usual. The conside  
approach until the final day of the

## Forecasting restaurants with critical violations in Chicago

Gene Leynes (City of Chicago), Anush Solanki (City of Chicago), Tim Schenk, Jr. (City of Chicago)

February 16, 2015

The Chicago Department of Public Health (CDPH) inspects more than 15,000 restaurants with fewer than three dozen inspectors over the course of the year. This paper describes a predictive model designed to identify the presence of a critical violation in a particular food establishment. The goal of this model is to prioritize inspections by likelihood in order to identify the riskiest restaurants earlier, thereby reducing the length of exposure of many restaurants to patrons. Critical violations were identified approximately 7.44 days earlier over a 60-day period compared to current operations in the out-of-sample test.

### 1 Introduction

In 2014 the Chicago Department of Public Health inspected performed over 20,000 inspections at nearly 13,000 food establishments across Chicago with fewer than three dozen inspectors. The majority of these food inspections were routine inspections that don't uncover serious problems, but some of these inspections uncovered issues that affect the health and safety of the patrons who visit these establishments. Traditionally, prioritizing these inspections is a largely manual task that relies on a combination of administrative processes and personal expertise.

The model set forth in this paper can help with the prioritization of scheduled, saving time and money as well as making the city's food safer. The model utilizes several data sources and through advanced modeling techniques the model provides additional insight into an establishment's current actual risk based on real-time data.

This paper is organized as follows: Section 1 provides an introduction and background to describe the current process and scope of the problem. Section 2 describes data that has been collected by the research team for this project and how that data was combined. Section 3 describes the model. Section 4 describes the model evaluation. Section 5 contains details of the model results from the experiment. Finally, the Summary section concludes with a brief summary of results and information regarding the ongoing project.

Ultimately, we find that a data-driven model can help inspectors discover critical violations earlier than the current "Business As Usual" (BAU) process. On average, critical violations would have been discovered 7.44 days earlier over the two-month test period. The first half of the experiment yielded 25.0% higher successful inspections. Beginning in 2015, CDPH has begun to use this analytical model to prioritize current inspections. Risk 1 and risk 2 Food Establishments will still undergo annual inspections; however, those restaurants with the highest likelihood of the most serious issues will be prioritized.

It is worth noting that this research is an open source project. The source code of the statistical model is available on the City of Chicago food inspection project page. The statistical modeling was completed using the open source statistical software R, and all the necessary data to replicate those results is available online. This paper was generated using Jupyter, which allows others to view the underlying calculations to generate the summaries, tables, and diagrams in this document. This document is available in the same aforementioned repository.

CDPH mainly conducts three different types of inspections: regular routine visits are the only type of inspection included in the model.

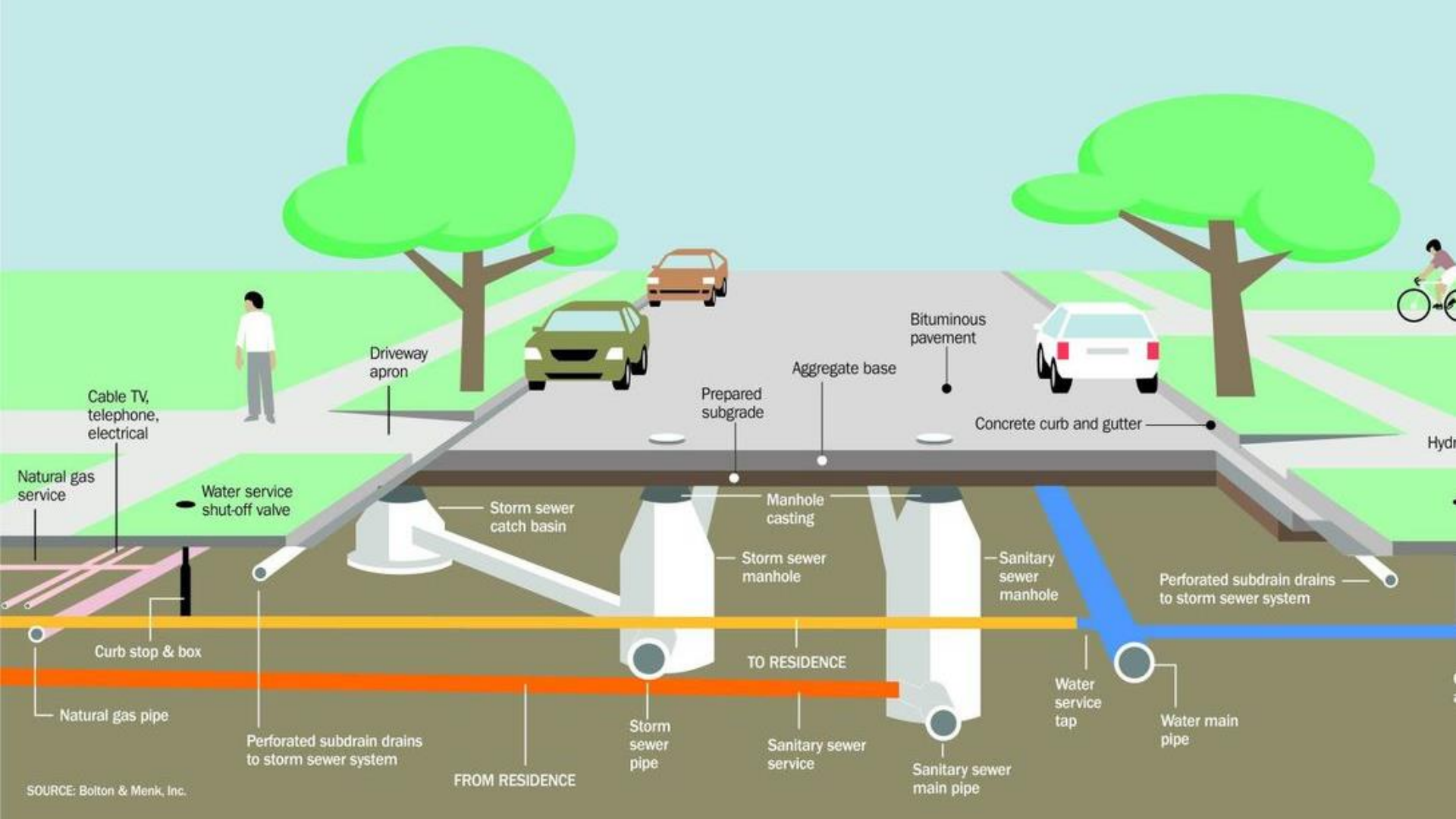
# Reproducible Research

The technical paper was written as a highly-reproducible “knitr” document, allowing other researchers to understand how summary numbers were calculated. Each statement in the project can be traced to an original source.

```
File Edit Selection Find View Goto Tools Project Preferences Help
forecasting-restaurants-with-critical-vi... mongoexport -u tom -p DataGuy -h my... red_light_camera_violations_spxq-js37... Tax_Increment_Financing_TIF_Proj...
211 # Introduction
212
213 In 2014 the [Chicago Department of Public Health](http://www.cityofchicago.org/city/en/depts/cdph.html) inspected performe
fewer than three dozen inspectors. The majority of these food inspections were routine inspections that don't uncover seri
and safety of the patrons who visit these establishments. Traditionally, prioritizing these inspections is a largely manu
.
214
215 The model set forth in this paper can help with the prioritization of scheduled, saving time and money as well as making t
modeling techniques the model provides additional insight into an establishment's current actual risk based on real-time d
216
217 This paper is organized as follows: Section 1 provides an introduction and background to describe the current process and
research team for this project and how that data was combined. Section 3 describes the model. Section 4 describes the mod
Finally, the Summary section concludes with a brief summary of results and information regarding the ongoing project.
218
219 Ultimately, we find that a data-driven model can help inspectors discover critical violations earlier than the current "B
discovered "r sprintf("%0.2f", mean(time_diff))" days earlier over the two-month test period. The first half of the exper
"%") higher successful inspections. Beginning in 2015, CDPH has begun to use this analytical model to prioritize canvass i
inspections; however, these restaurants with the highest likelihood of the most serious issues will be prioritized.
220
221 It is worth noting that this research is an open source project. The source code of the statistical model is available on
inspections-evaluation). The statistical modeling was completed using the open source statistical software R, and all the
generated using [knitr](http://yihui.name/knitr), which allows others to view the underlying calculations to generate the
same aforementioned repository.
222
223 CDPH mainly conducts three different types of inspections; regular canvass inspections, new license inspections, and inspe
of inspection included in the model.
224
225 Before a food establishment opens their doors CDPH conducts an initial new license inspection. New license inspections are
cityofchicago.org/city/en/depts/bacp.html), who grants food establishment licenses to new establishments. Each establishme
Establishments often fail these initial inspections because they have not yet finished setting-up equipment, such as turni
CDPH will re-inspect those establishments to ensure those conditions are passed before they are allowed to open. New lic
characteristic of normal inspections, and because they occur when a new business applies for a license, and are therefore
226
227 The majority of the food inspections are regular canvass visits, which must be done on a regular basis to check the quali
level of the facility. Risk 1 establishments are ideally inspected two times a year.
228
229 If a restaurant fails the inspection, because of violations / citations, the inspector will return at a later date to see
and are not included in the model. Only the initial canvass inspection is included.
230
231 The third type of inspection occurs when complaints are registered from residents, alderman, and referrals from hospitals
be submitted through residents calling 311 or submitting a request through an online form. Individuals are asked to submit
describe the symptoms and what was eaten, and when it happened. CDPH reviews the materials and may initiate a food inspect
232
233 A breakdown of the inspection types in 2014:
234
235 ```{r}
236 kable(insp_types_2014[i = TRUE,
```



**UNDERGROUND  
INFRASTRUCTURE IS HIT ON  
AVERAGE EVERY 60 SECONDS.  
THE TOTAL COST TO THE  
NATIONAL ECONOMY IS  
ESTIMATED TO BE \$1.6 BILLION**





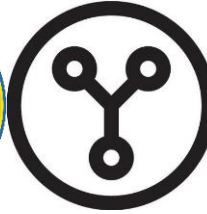


## Underground Map

Using off-the-shelf DSLR cameras, photos are stitched together to create a 3-D model of the city's underground infrastructure. City Digital, City of Chicago and a consortium of partners are piloting the tech.

# THANK YOU

---



## **Tom Schenk Jr.**

Chief Data Officer  
City of Chicago

## **Contact Info:**

@ChicagoCDO  
tom.schenk@cityofchicago.org

## **Websites:**

Data @ [data.cityofchicago.org](http://data.cityofchicago.org)  
News @ [digital.cityofchicago.org](http://digital.cityofchicago.org)  
Maps @ [opengrid.io](http://opengrid.io)  
Strategy @ [techplan.cityofchicago.org](http://techplan.cityofchicago.org)  
Slides @ [speakerdeck.com/tomschenkjr](http://speakerdeck.com/tomschenkjr)