REAL WORLD R

USING R IN THE REAL WORLD
STAT 385: STATISTICS PROGRAMMING METHODS
DECEMBER 11, 2017

About Gene

- DePaul University – Math / Stats / Actuarial
- Milliman – Casualty Actuarial Consultant
  - Non traditional clients
  - Great learning experience
  - Worked on over 100 projects over 6 years
- University of Chicago MBA
- Entrepreneur in algorithmic trading / finance
- Entrepreneur in consulting – Chicago Data Science
- City of Chicago – Department of Innovation and Technology

City of Chicago

Employees by Department

CITY COUNCIL	370
BOARD OF ELECTION	114
MAYOR'S OFFICE	96
PROCUREMENT	79
HUMAN RESOURCES	68
INSPECTOR GENERAL	55
BUDGET & MGMT	43
DISABILITIES	29
TREASURER	24
HUMAN RELATIONS	18
BOARD OF ETHICS	9

FAMILY & SUPPORT	726

PUBLIC HEALTH	578

GENERAL SERVICES	930
IPRA	87

POLICE	13,682
FIRE	4,706
OEMC	1,241
POLICE BOARD	2

STREETS & SAN	2,222

WATER MGMNT	1,830

AVIATION	1,474

TRANSPORTATION	1,030

PUBLIC LIBRARY	993

FINANCE	545

LAW	459
ADMIN HEARNG	41

BUILDINGS	256

COMMUNITY DEVELOPMENT	218

BUSINESS AFFAIRS	182

DoIT	96

CULTURAL AFFAIRS	75

CITY CLERK	87

ANIMAL CONTRL	66

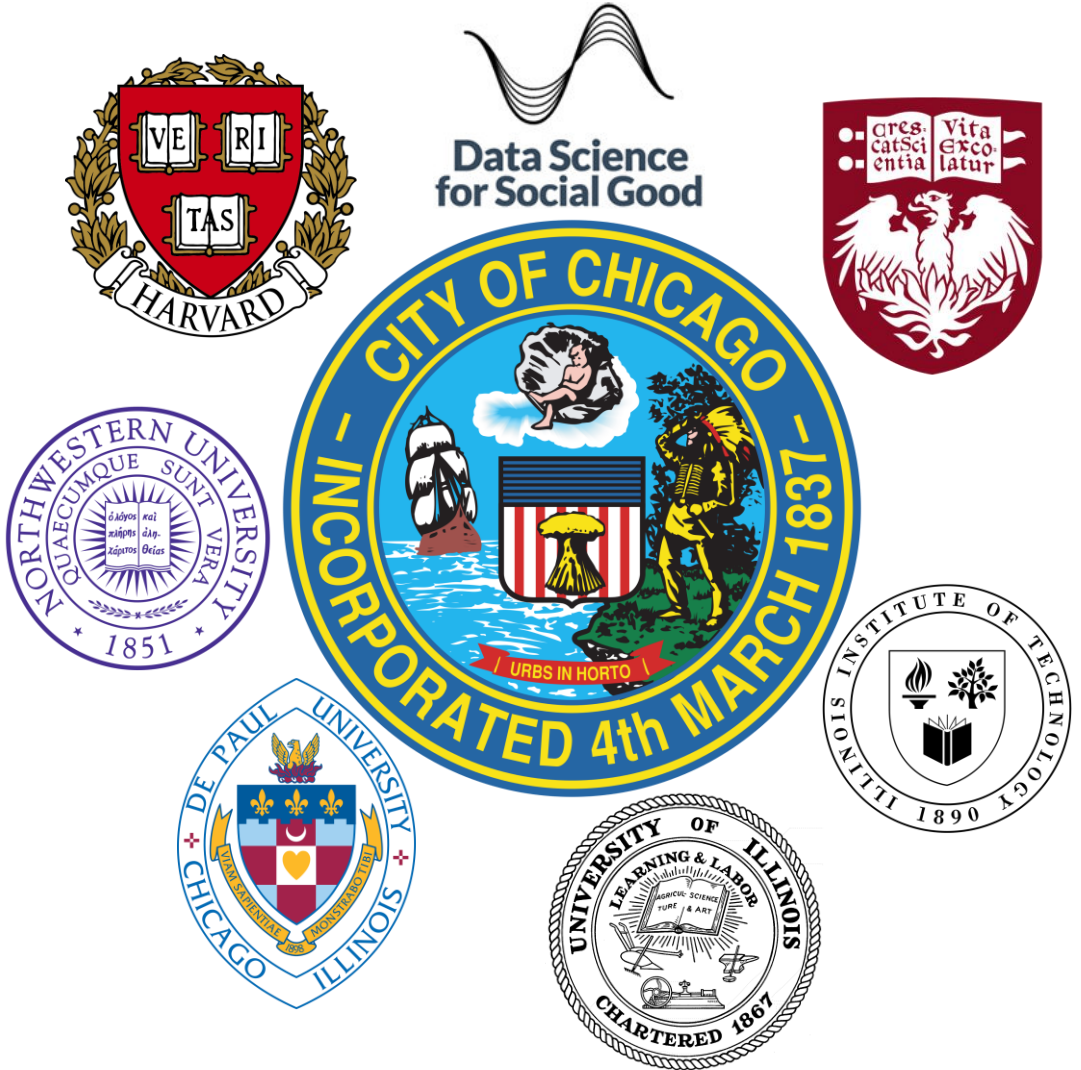
Sister Agencies

cta

Chicago Public Schools

Research Partners

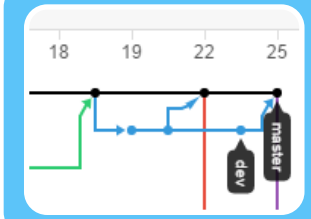
The City of Chicago has a number of high-quality research universities and groups willing to engage in projects with the city. We can leverage open data portal and data itself to create cooperative relationships.



Successful Technical Project

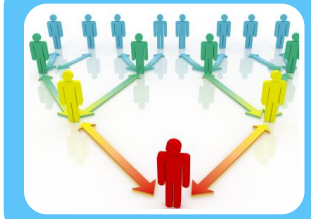
Good Client

- Wants the answer
- Available
- Engaged (expectation management)



Workflow

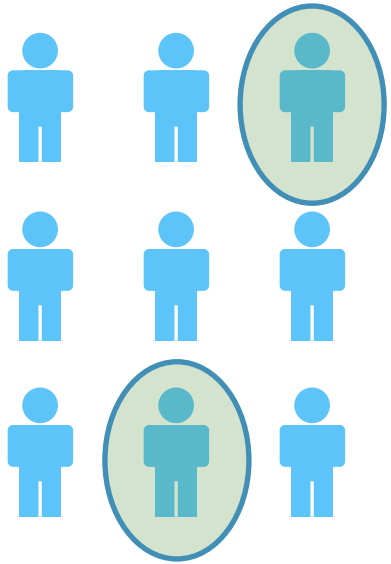
- Well defined problem
- GitHub or similar to manage issues and communication
- Reproducible

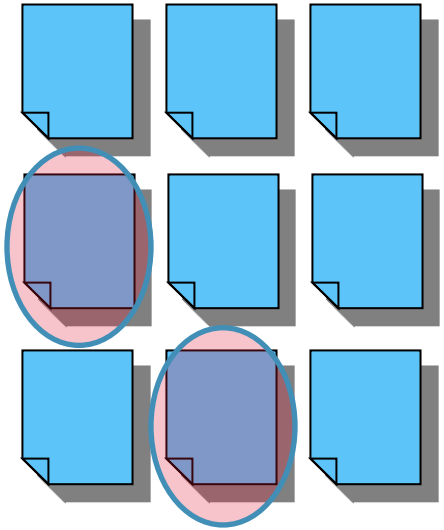


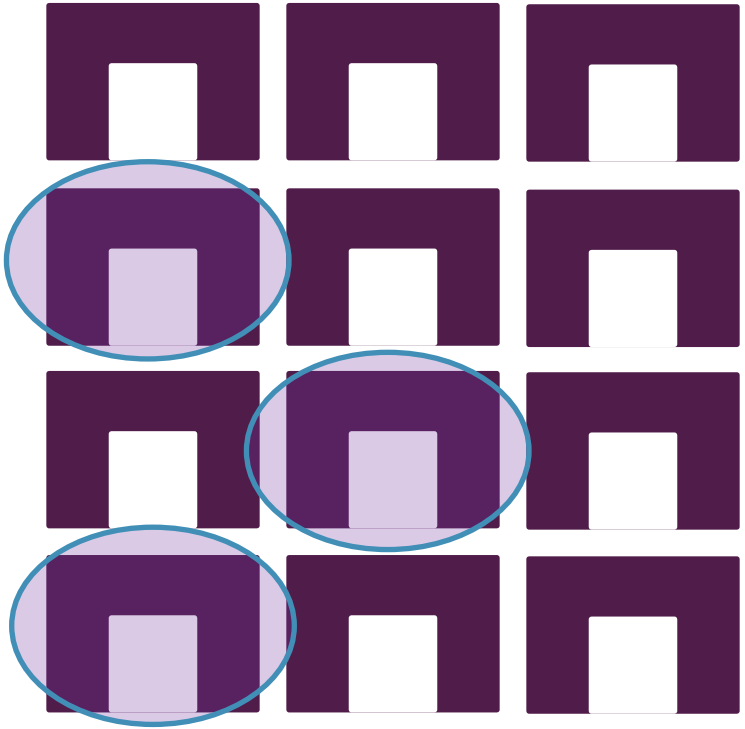
Internal Support

- Organizational alignment
- Good team
- Consistent data

Identify “y”



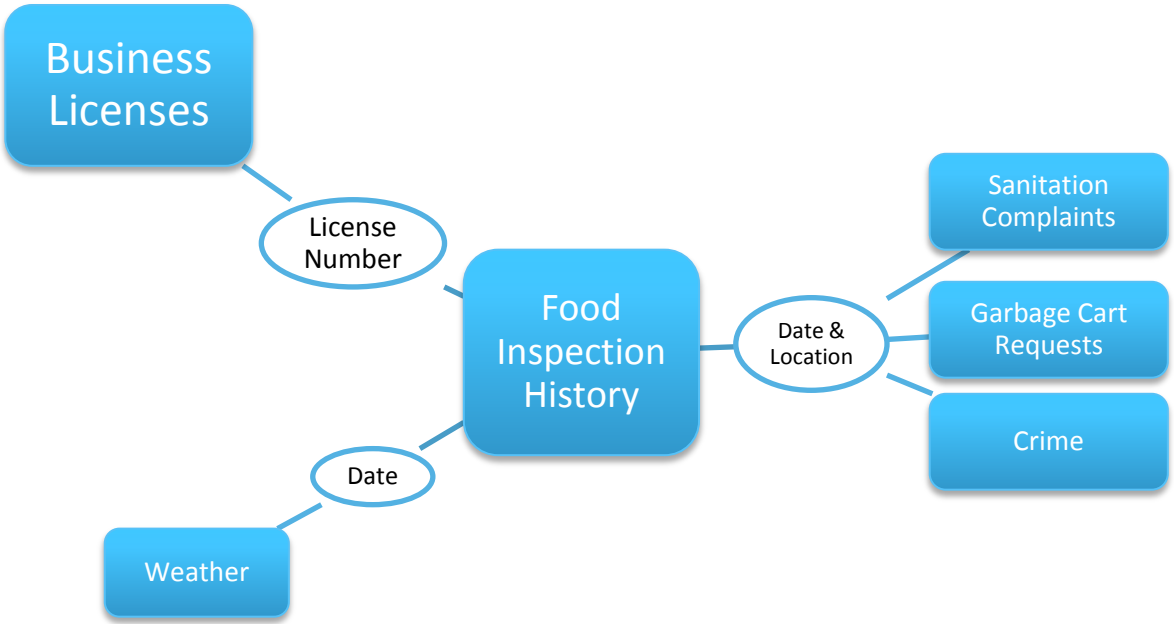




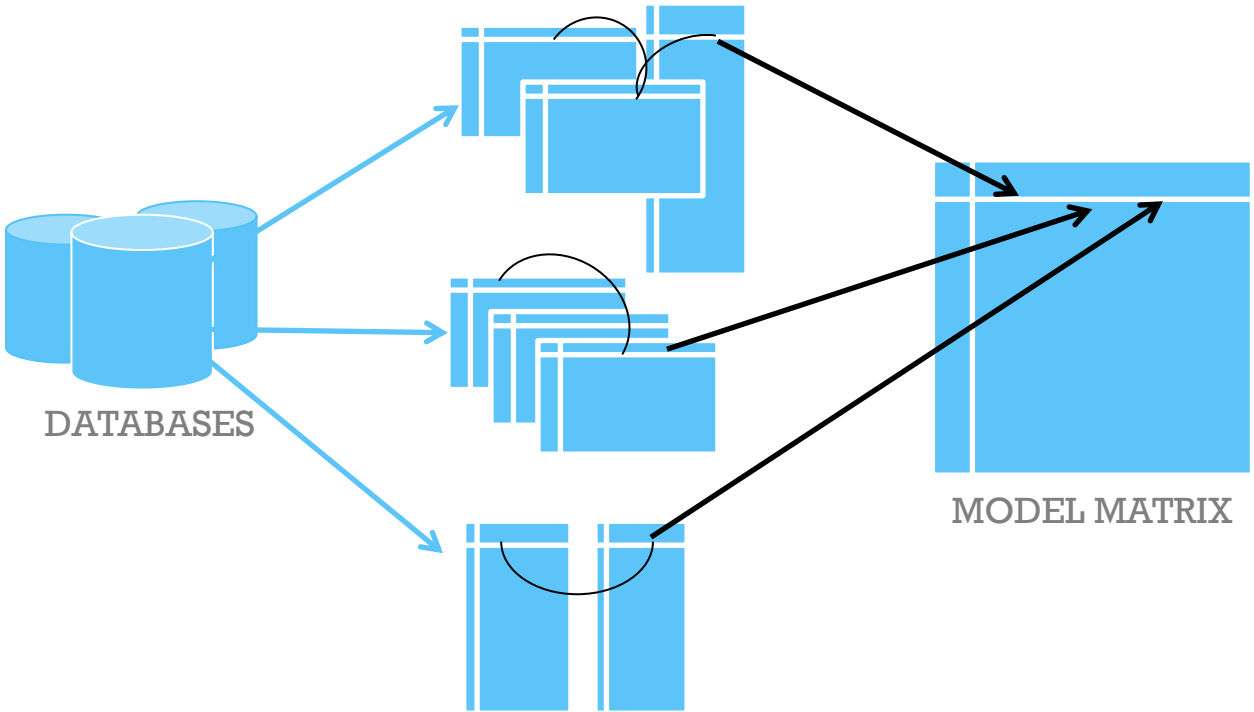
Defining the response variable (y) drives model development

- Each row in x, (the model matrix) is connected to a y
- It should be possible to validate y against other reports
- The client / user may not initially realize what they want to predict
  - The definition of y might change
  - The outcome of one y might be the input to another model

Example: Food Inspection Data



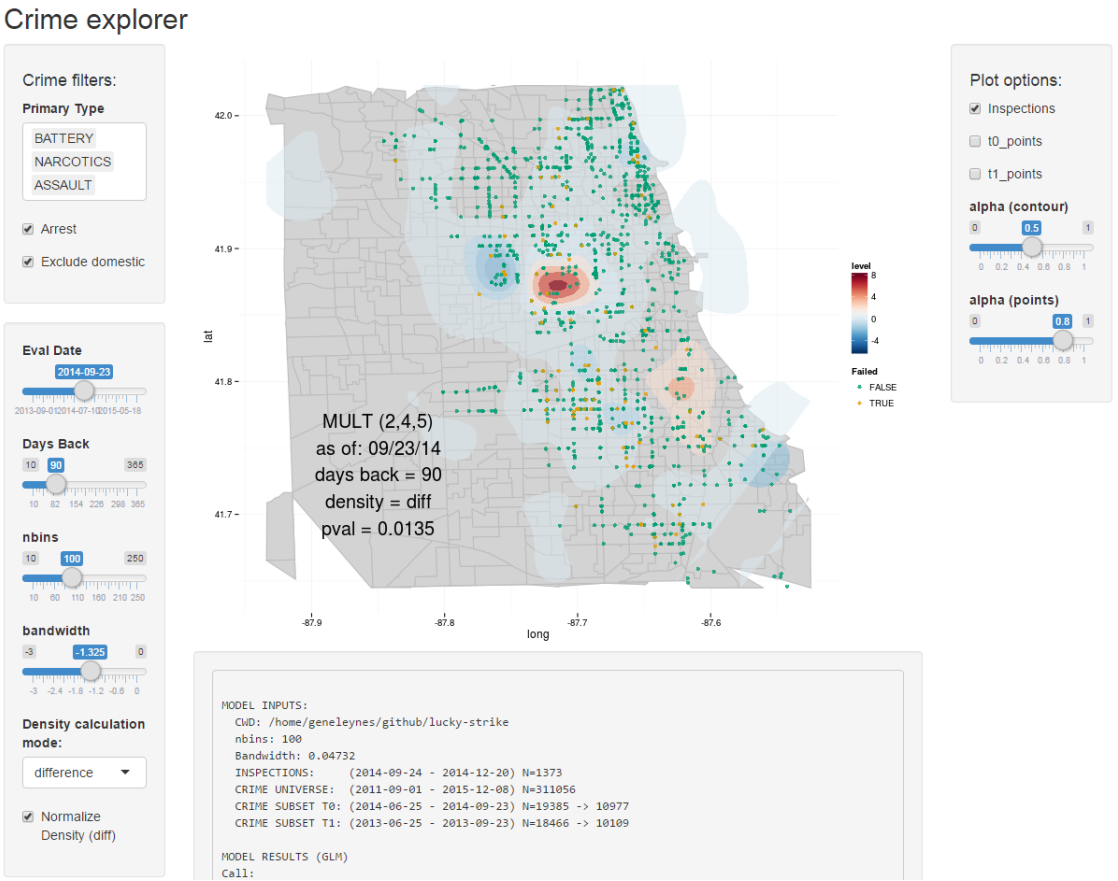
Feature Engineering: Calculate “x”



Main challenges in calculating features:

- Finding relationships between tables in databases
- Consolidating historical features to one record

Match x to y – by date and location

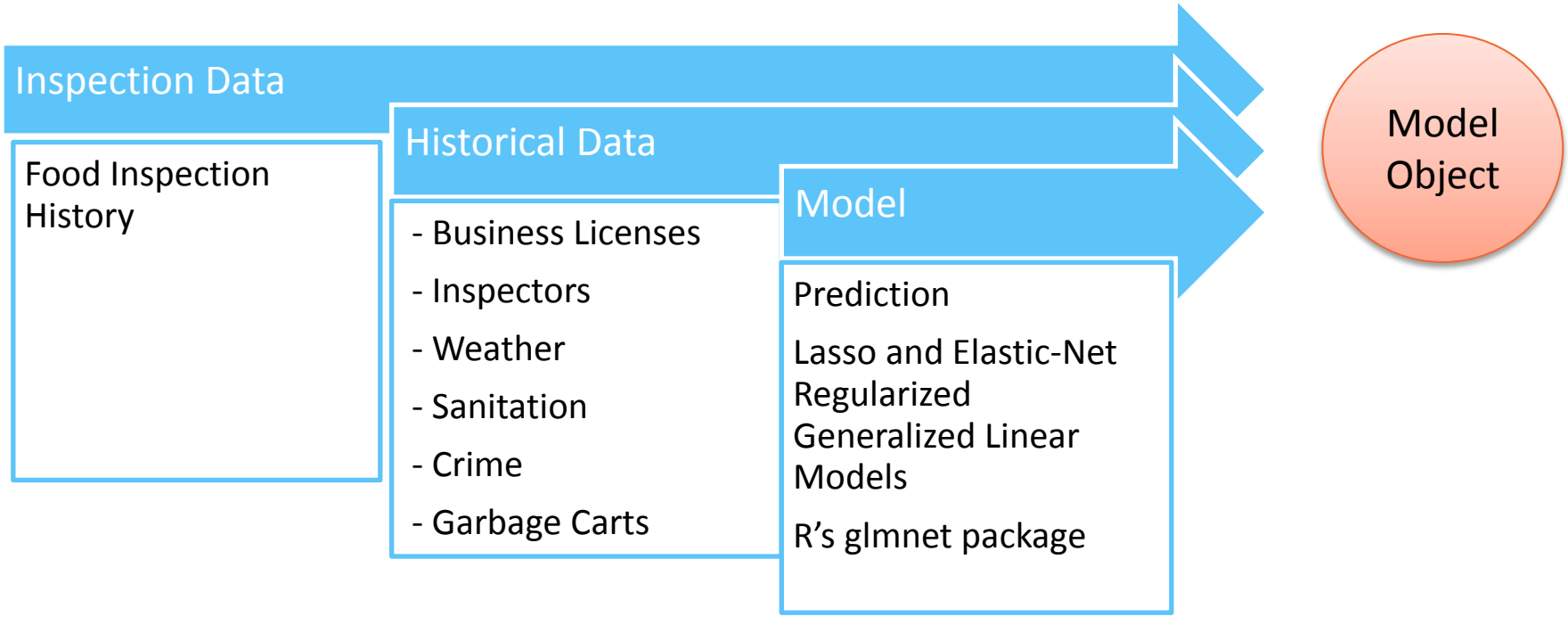


311 & 911 phone calls come with a date and a location. We can connect them to the date and location of inspections as follows:

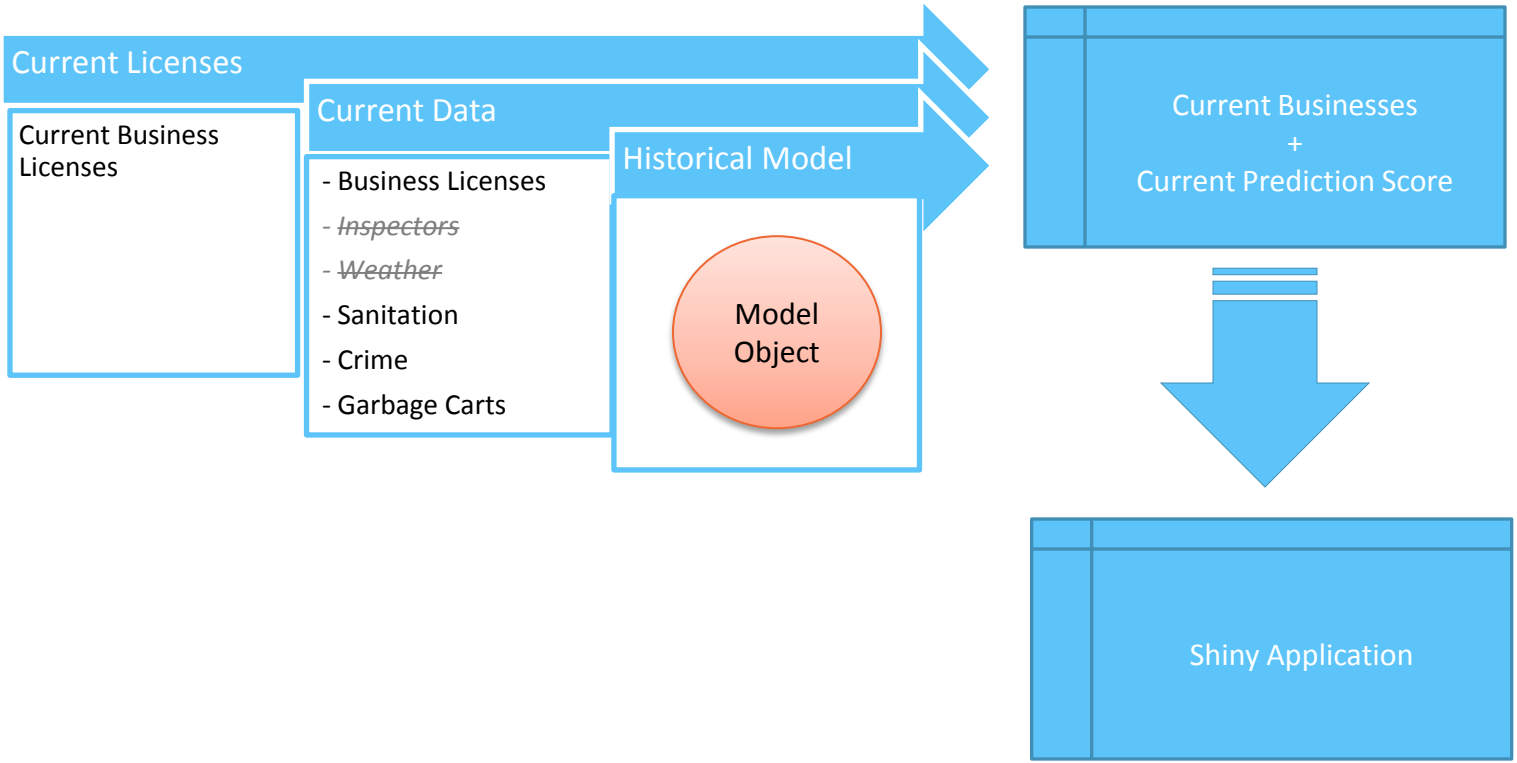
- Choose types of calls to include in analysis
- How many days to look back
- Choose whether to look at absolute levels, or year over year differences
- Choose a bandwidth
- Test relationship over various time horizons

This is not easy!

Model


Inspection Data
Food Inspection History
Historical Data
- Business Licenses
- Inspectors
- Weather
- Sanitation
- Crime
- Garbage Carts
Model
Prediction
Lasso and Elastic-Net Regularized Generalized Linear Models
R's glmnet package
Model Object

Prediction and Application

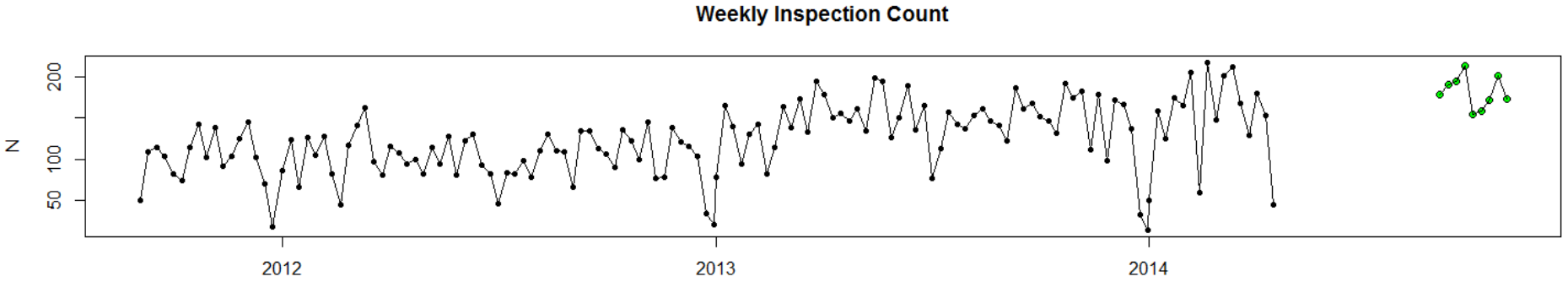


The Test / Train Framework

- The first model was built using data prior to 2014, and tested in early 2014
- The second model was built in response to the first
  - Completed in the summer of 2014
  - Tested in November based on actual inspection results from September and October

Sept 2011 – Feb 2014

Sept 2014 – Oct 2014



Shiny Application

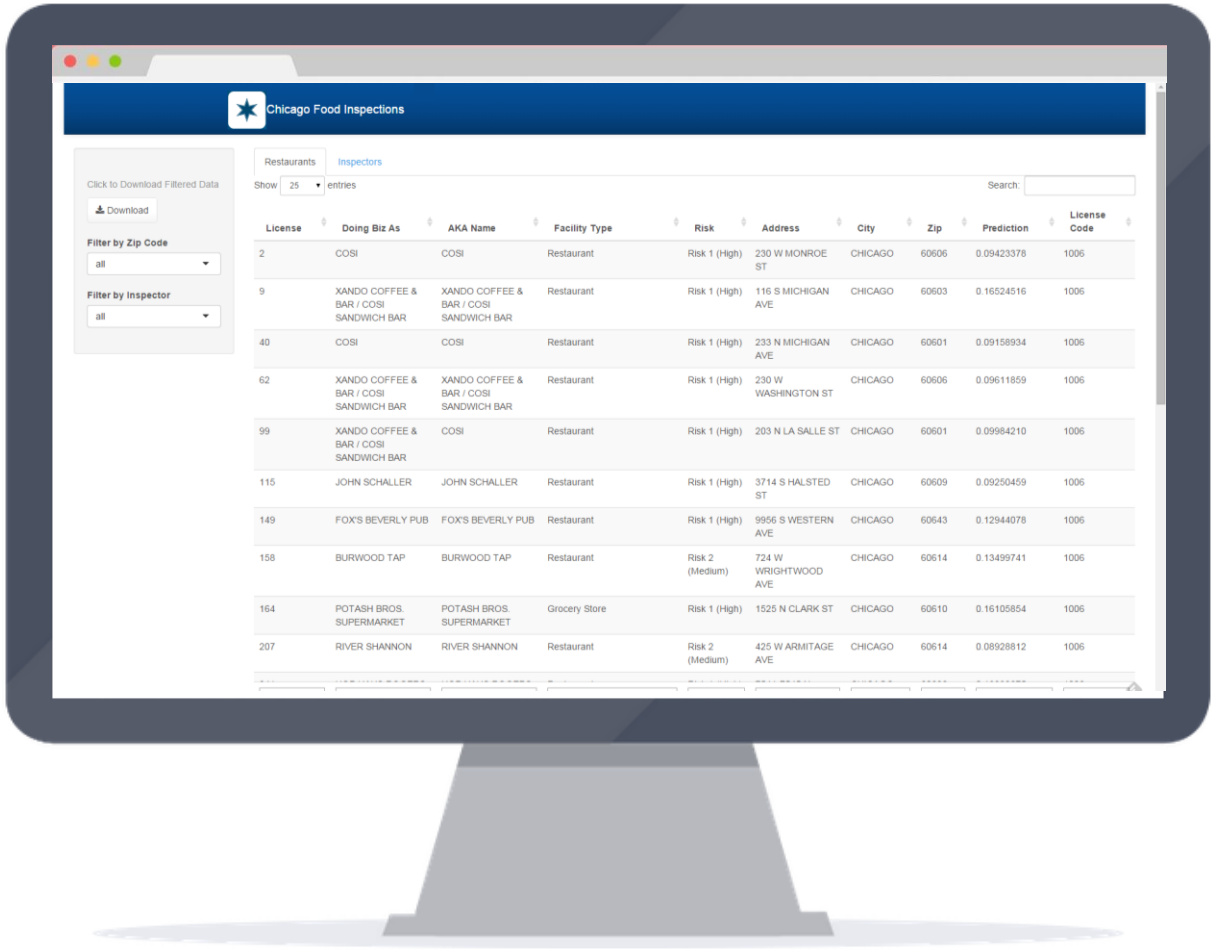
- Based on simple framework, with only three files
  - global.R
    - Sets global data variables (imports pre-calculated data, renames fields)
  - sever.R

```
shinyServer(
  function(input, output) {
```

  - ui.R

```
shinyUI(
  fluidPage(
    <header elements>
  navbarPage(
    <page elements>
  tabPanel(
    <tab elements>
  <etc…>))))
```

References objects created in server.R



The Application

Our final outcome was a simple list that contained

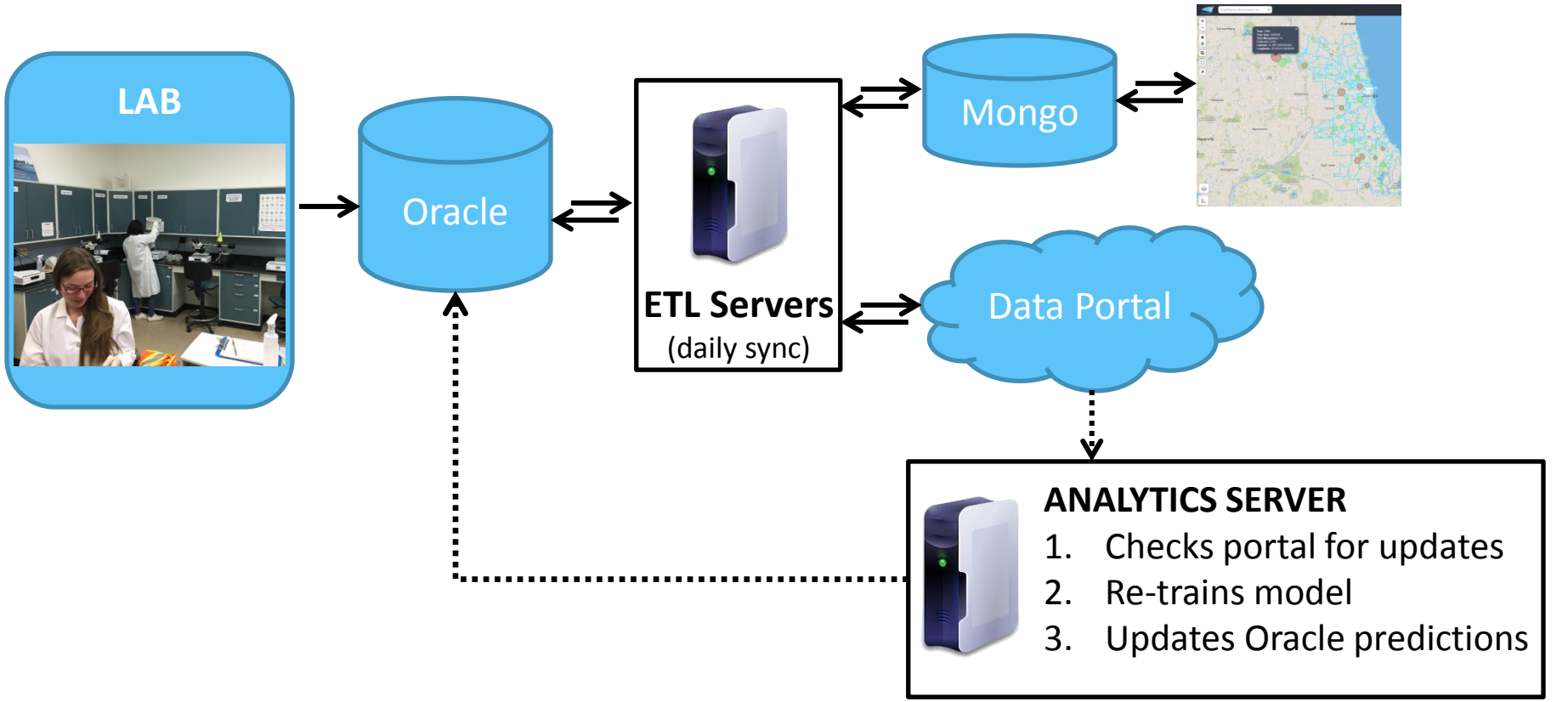
- Business details
- Zip codes
- Predictions

That's it, no fancy maps!

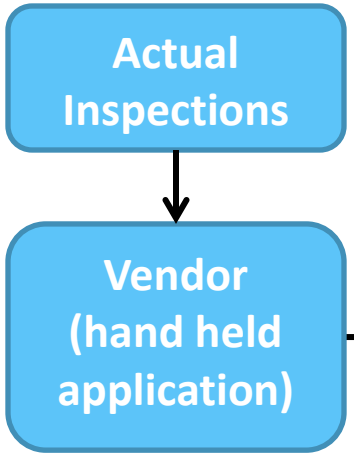
Technical notes:

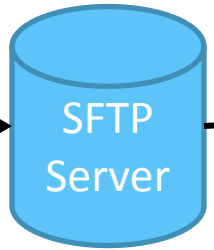
- Updates nightly via shell script (**crontab**!!!)
- Uses R Studio's Shiny
- MVC framework
- Built on JQuery

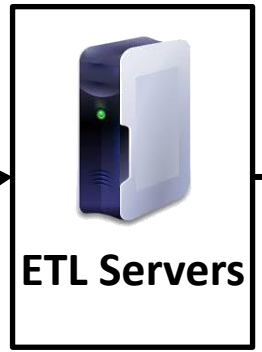
Workflow Example – West Nile Virus

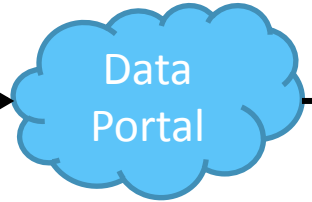


Workflow Example – Food Inspections



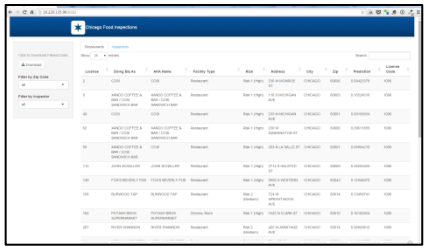






ANALYTICS SERVER

1. Pulls new data nightly
2. Re-trains model
3. Hosts predictions in Shiny app



THANK YOU

Gene Leynes
Data Scientist
City of Chicago

Contact Info:

@geneorama
gene.leynes@cityofchicago.org

Websites:

cityofchicago.org
techplan.cityofchicago.org
data.cityofchicago.org

github.com/Chicago
github.com/Chicago/food-inspections-evaluation
github.com/Chicago/west-nile-virus-predictions