Data Science Through the Lens of Social Science

Drew Conway
Head of Data, Project Florida
Co-founder, DataKind

Project Florida

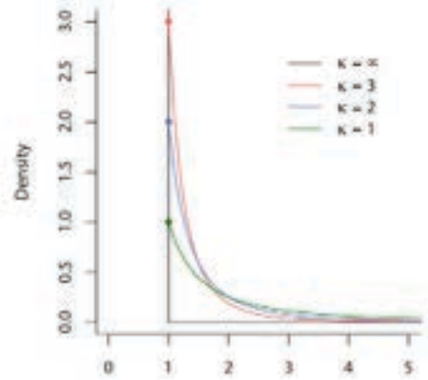
DataKind
USING DATA IN THE SERVICE OF HUMANITY

@drewconway

I want to tell you a story…

Kelly embarks on a *data science journey*!

…she consumes math, stats, and CS material



…she spends free time hacking on side projects

…she graduates from a top university with

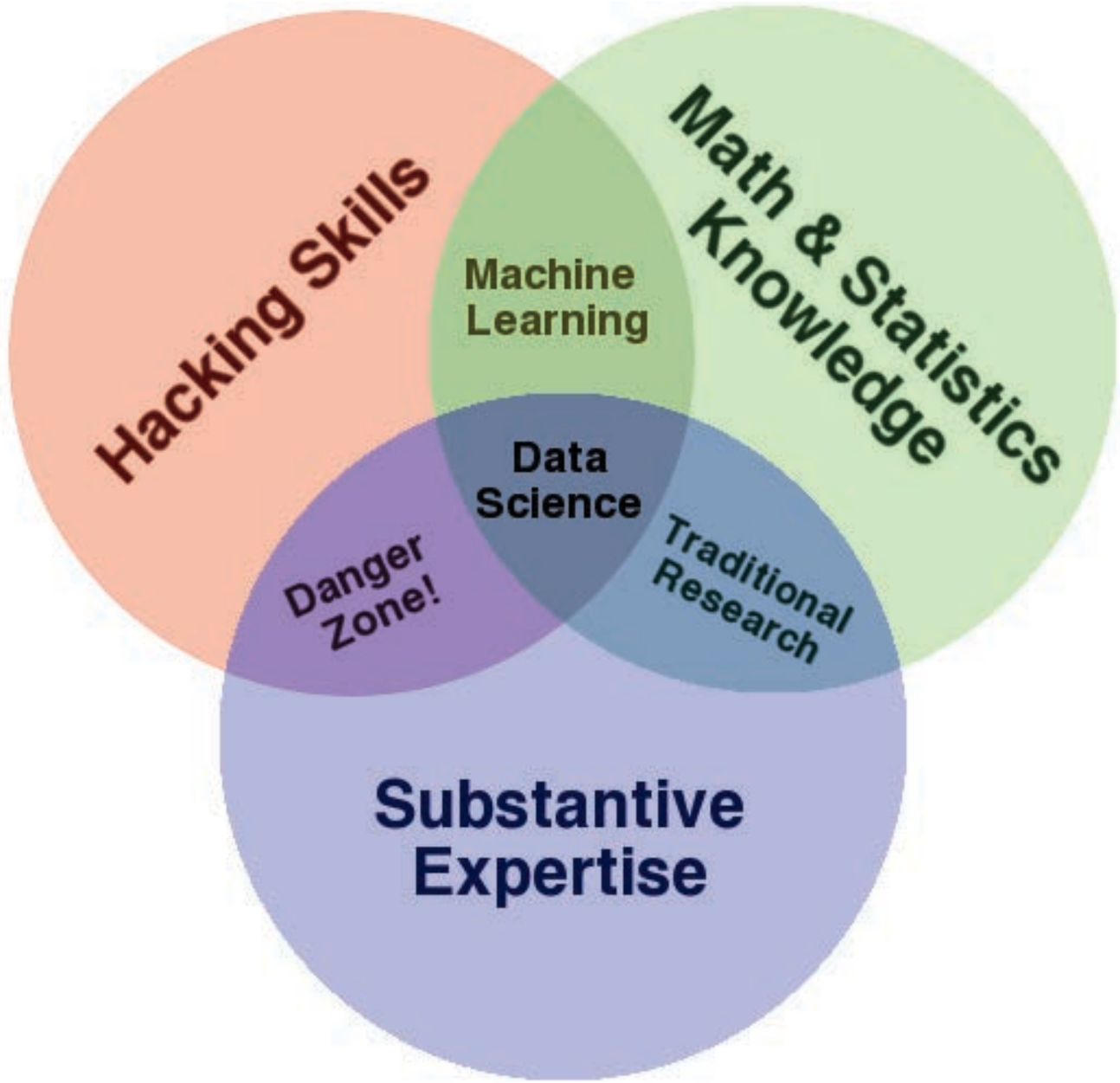
…she lands a lucrative job at an ad-tech firm!

No young hacker grows up dreaming of selling ads!

So, why do so many end up doing it?

Say “big data” again!

Source: http://bigdatapix.tumblr.com/


Hacking Skills
Math & Statistics Knowledge
Machine Learning
Data Science
Danger Zone!
Traditional Research
Substantive Expertise

Math & Stats Knowledge

If: Better data beats better math

Then: What methods should be taught?

Explore

How do you find structure in new data?

- Scatter plots
- Density plots

Data exploration that scales

- Reduce dimensionality
- Univariate analysis

Model

Methods must match data

- Text
- Geospatial
- Web-scale

What is the 'best' model?

- Most predictive
- Most parsimonious
- Cross-validation

Math & Stats Knowledge

Universities good at methods training...
...but what methods fit into Data Science?

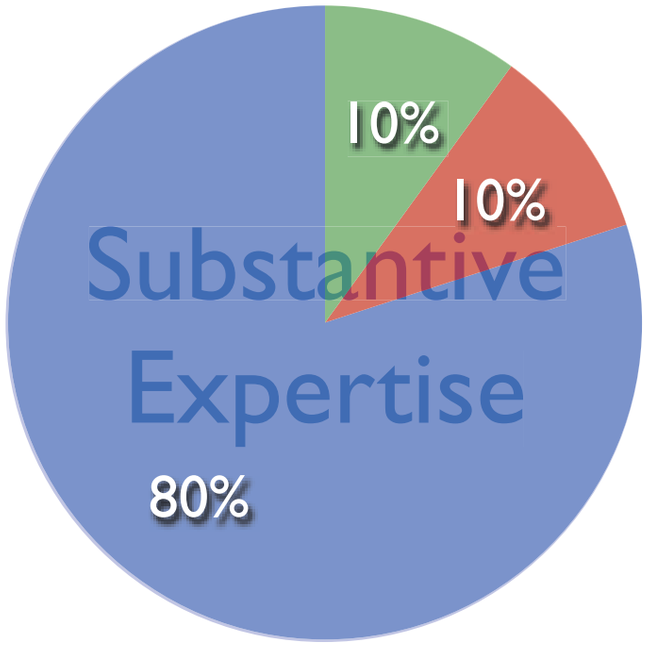
Things data scientist like...

- Illustrating the current state of the world
- Predicting future observations
- Classifying/ranking observations

Things social scientists like...

- Testable theoretical models
- Natural experiments
- Causality

1. When applicable
2. Right tool / right job
3. Open black boxes
4. Learn limitations



Data Science, as a discipline, is fundamentally about **human behavior**

`Inquire`

`Interpret`

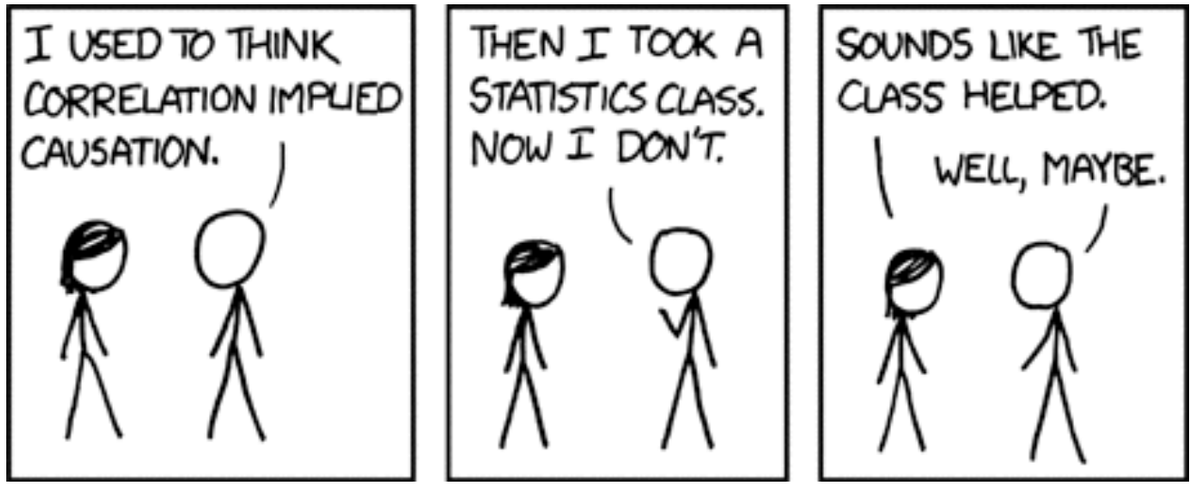
Focus on questions / not tech

- What new questions can be asked from web-scale data?
- Tools are a means to an end

Social science has questions

- Markets
- Organization
- Decision making

How do we know when the results we get make sense, if ever?



https://flic.kr/p/7zaJyV

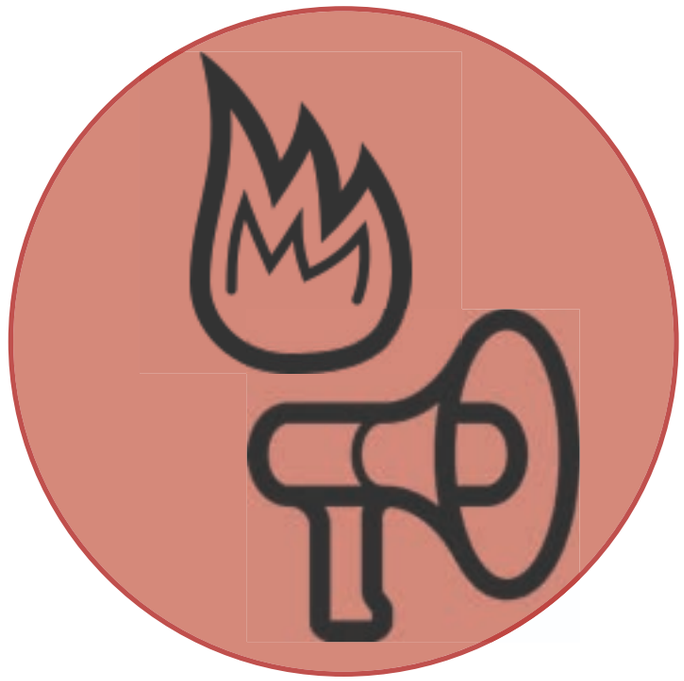
NYC Analytics

MODA Analytics Projects

Day-to-day operations

Economic development

Disaster response & sustainability



Representative NYC agencies

- Dep't of Buildings/HPD
- Sherriff's Office
- Dep't of Environmental Protection
- Dep't of Sanitation
- NYPD/FDNY

- Economic Development Corporation
- New Business Acceleration Team
- Small Business Services
- Dep't of Finance

- Housing Recovery Office
- Office of Long-Term Planning & Sustainability
- Office of Emergency Management

SUPERVISING
FIRE MARSHAL
FDNY

Day-to-day operations: FDNY RBIS

(Risk Based Inspection System)

ROC curve pre-analysis

ROC curve post-analysis

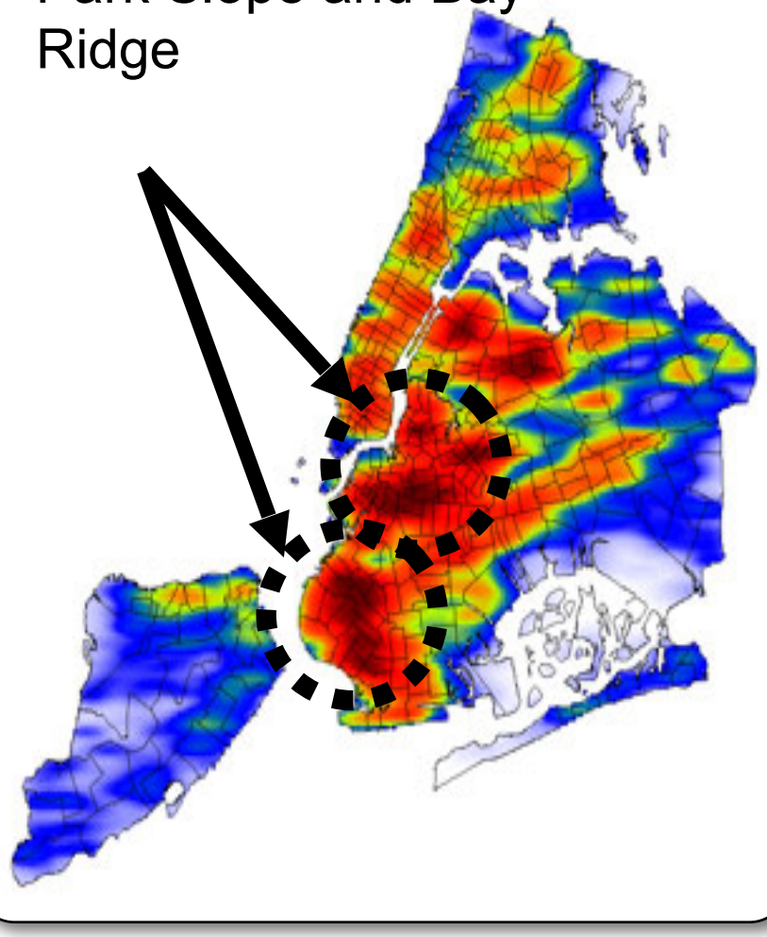
Re-ranking inspections significantly reduces the time that buildings are at risk of severe fire

Day-to-day operations: FDNY RBIS

Improved results from RBIS update

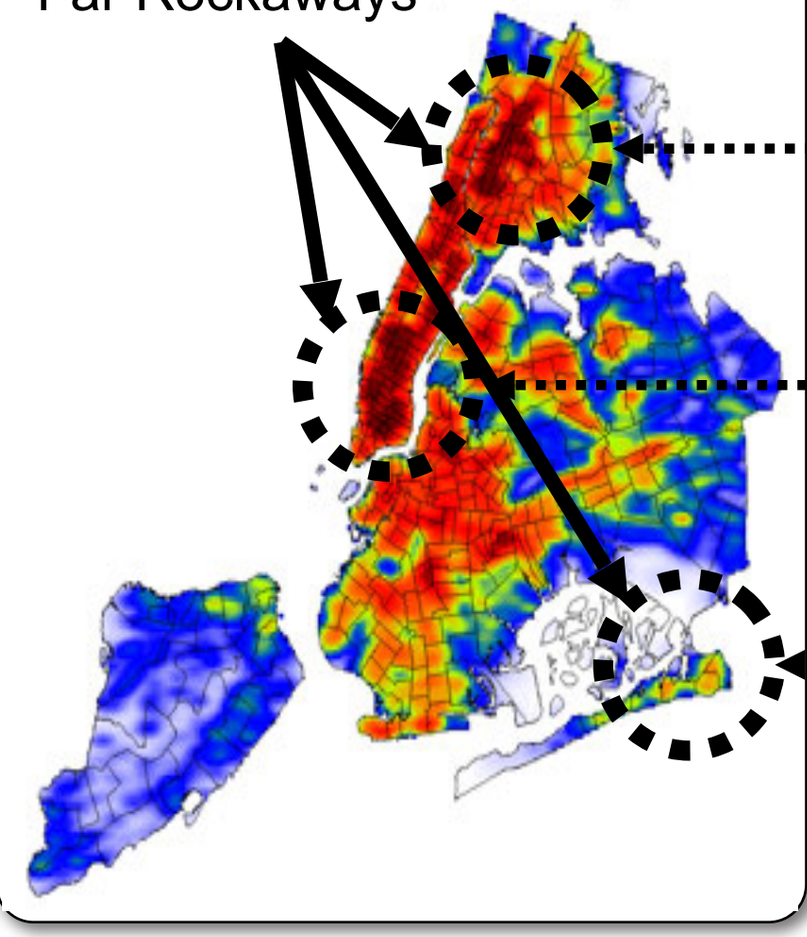
Original model

Original model over-predicted fires in Downtown Brooklyn, Park Slope and Bay Ridge



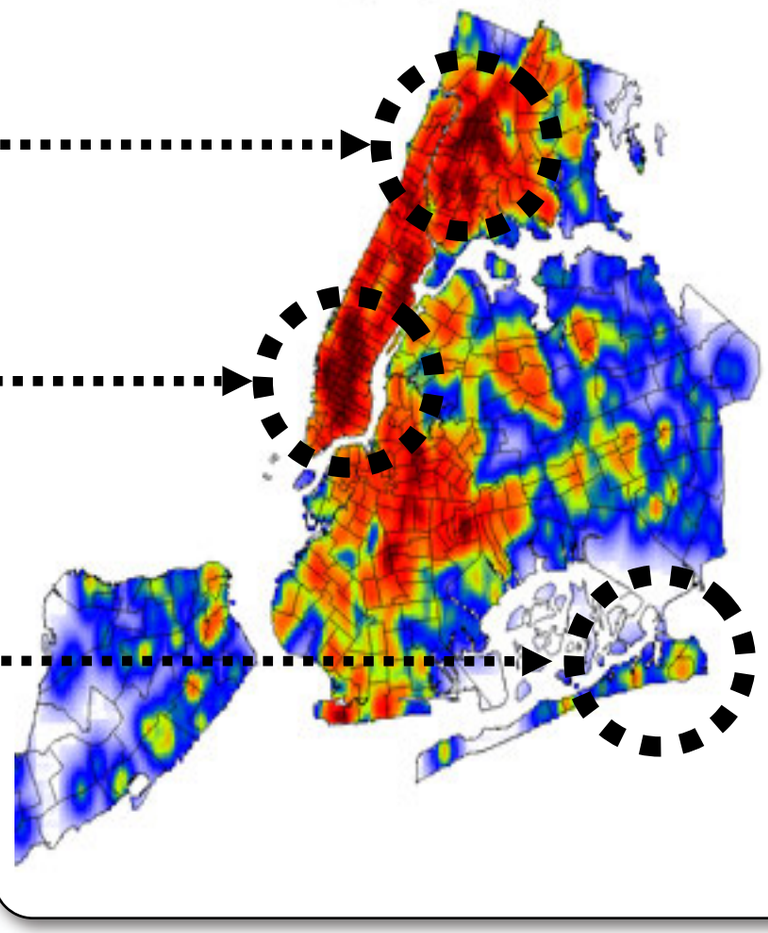
Updated model

Updated model accurately reflects risk in West Bronx, Downtown, and Far Rockaways



Actual fires

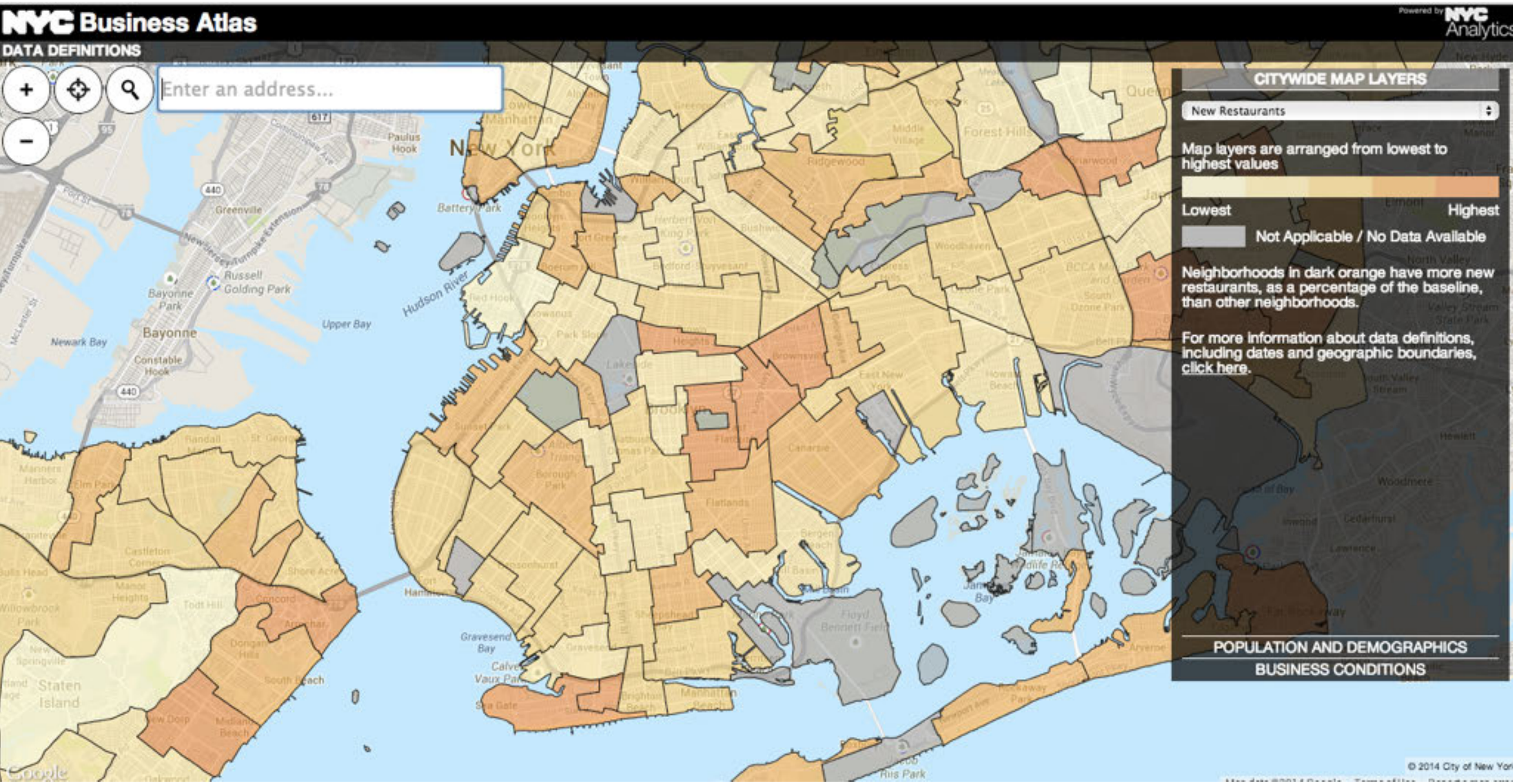
Observed fire frequency, 2011 to present



Come in
WE'RE
OPEN
SOURCE

Business Atlas

Helping entrepreneurs make smart decisions



http://maps.nyc.gov/businessatlas/

ON THE CLOCK

NEW METHOD

9 MIN. 42 SEC.

From when 911 is reached

OLD METHOD

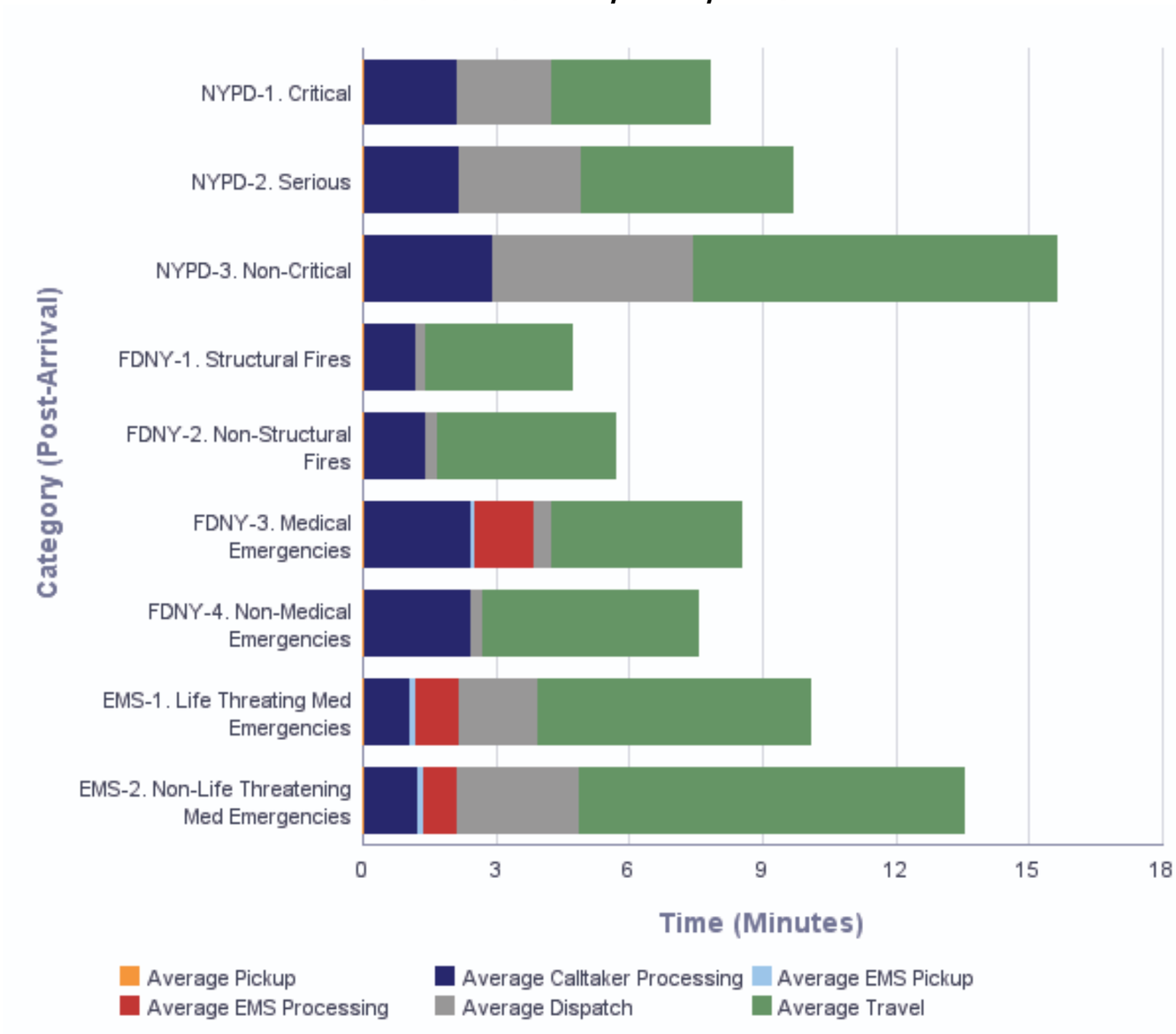
6 MIN. 47 SEC.

From when call transferred to appropriate dispatcher

Source: NY Daily News

911 end-to-end
www.nyc.gov/911reporting

Week of 7/21/2014



Pharmacy

Project Florida

drew@projectfla.com