## How to Change a City with Data

Ben Wellington

Two Sigma Investments, I Quant NY

iquant.nyc @iquantny

# Data Storyteller

## Data Science

## Data Science

## Urban Planning



"Across City government, agencies use data to develop policy, implement programs, and track performance — and each month, our Administration shares more and more of this data with the public at large, catalyzing the creativity, intellect, and enterprising spirit of computer programmers to build tools that help us all improve our lives."













SIGN UP FOR UPDATES















## Work together to solve New York's challenges.

BigApps is a competition that empowers the sharpest minds in tech, design, and business to solve NYC's toughest challenges. From May to July 2014, over 100 teams built, pitched, and demonstrated strong and innovative projects. 20 BigApps finalists are moving on to compete for over \$100,000 in cash prizes!

**SEE GRAND PRIZE WINNERS** 

**SEE BIGIDEA WINNERS** 

SEE ALL FINALISTS

"Across City government, agencies use data to develop policy, implement programs, and track performance — and each month, our Administration shares more and more of this data with the public at large, catalyzing the creativity, intellect, and enterprising spirit of computer programmers to build tools that help us all improve our lives."

"Across City government, agencies use data to develop policy, implement programs, and track performance — and each month, our Administration shares more and more of this data with the public at large, catalyzing the creativity, intellect, and enterprising spirit of computer programmers to build tools that help us all improve our lives.

Open Data

Computer Programmers





**Open Data** 

Computer Programmers

Apps

Public

Policy Makers, Non-profits, Planners, News Agencies, Bloggers, Citizens, Advocates, etc.

#### Click here for the official list of NYC datasets



Business



City Government



Education



Environment



Health



Housing & Development



Public Safety



Recreation



Social Services

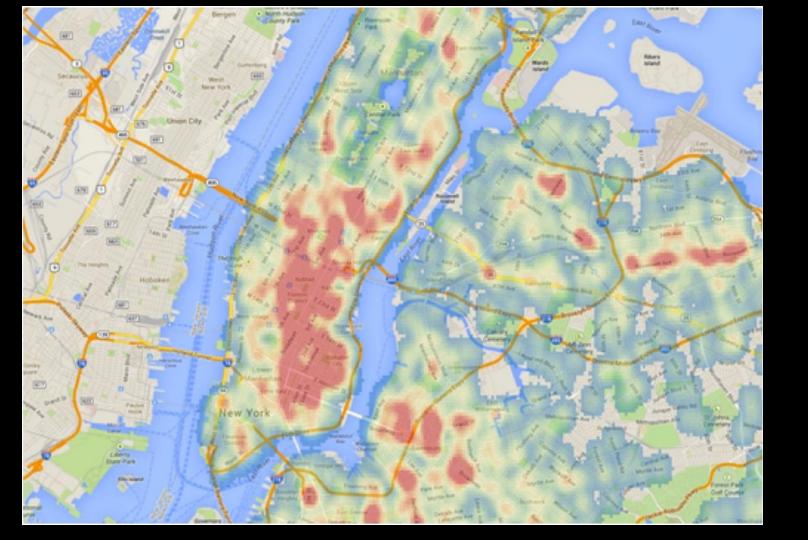


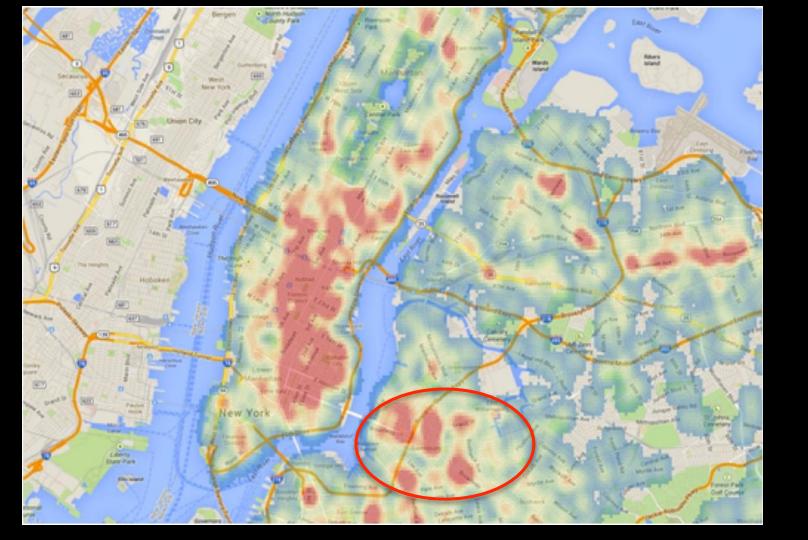
Transportation

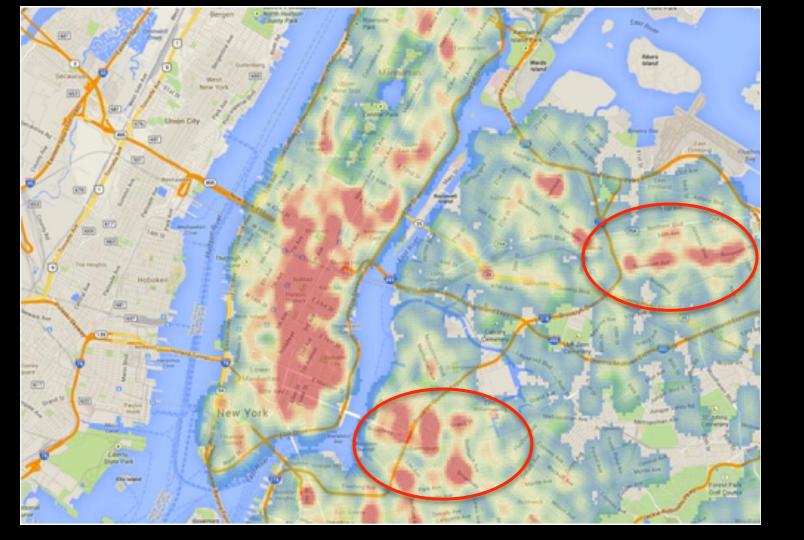


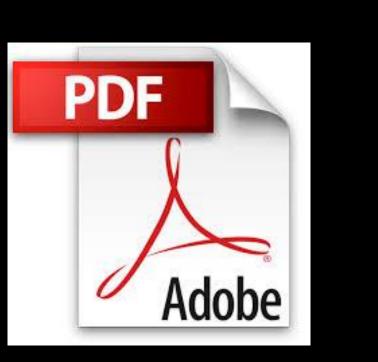
NYC BigApps











### Police Department City of New York

### sicle Collision Report Inte

UEENS					gust 20		Intersections		NYPD Precinc	ts Ma	
Intersection Address (***)	Number of Collisions	Persons Involved(*)	Collisions with Injuries	Category	Injured	Killed	Vehicle Types		Contributing Factors (**)		
100th Precinct											
ALMEDA AVENUE and BEACH 64 STREET	1	2	0	Motorists Passengers Cyclists Pedestr Total	0 0 0 0	0 0 0 0 0	Passenger vehicle	1	Driver inattention/distraction	2	
							Sport utility / station wagon	1			
BEACH 104 STREET and BEACH CHANNEL	1	2	0	Motorists Passengers Cyclists Pedestr Total	0 0 0 0	0 0 0 0 0	Large com veh(6 or more tires)	1			
DRIVE							Passenger vehicle	1			
BEACH 108 STREET	3	5	2	Motorists Passengers Cyclists Pedestr Total	3 0 0 0 3	0 0 0 0	Motorcycle	1	Failure to yield right-of-	1	
and BEACH CHANNEL DRIVE							Passenger 1		way		
							vehicle		Unsafe lane changing	1	
							Sport utility / station wagon	2			
							Unknown	1			
	1	1	1	Motorists Passengers	1 0	0	Motorcycle	1			
				Cyclists	0	0					

### **Police Department** City of New York

#### **Motor Vehicle Collision Report Intersections** August 2014

OUEENS		August 2014						NYPD Precincts I			
Intersection Address (***)	Number of Collisions	Persons Involved(*)	Collisions with Injuries	Category	Injured	Killed	Vehicle Types		Contributing Factors (**)		
100th Precinct											
ALMEDA AVENUE and BEACH 64 STREET	1	2	0	Motorists Passengers Cyclists Pedestr Total	0 0 0 0	0 0 0 0	Passenger vehicle	1	Driver inattention/distraction	2	
							Sport utility / station wagon	1			
BEACH 104 STREET and BEACH CHANNEL DRIVE	1	2	0	Motorists Passengers Cyclists Pedestr Total	0 0 0 0	0 0 0 0 0	Large com veh(6 or more tires)	1			
							Passenger vehicle	1			
BEACH 108 STREET and BEACH CHANNEL DRIVE	3 5	5	2	Motorists Passengers Cyclists Pedestr Total	3 0 0 0 3	0 0 0 0	Motorcycle	1	Failure to yield right-of-		
							Passenger 1 vehicle	way			
								200	Unsafe lane changing	1	
							Sport utility / station wagon	2			
						Unknown	1				

Motorists

Passengers Cyclists

Motorcycle

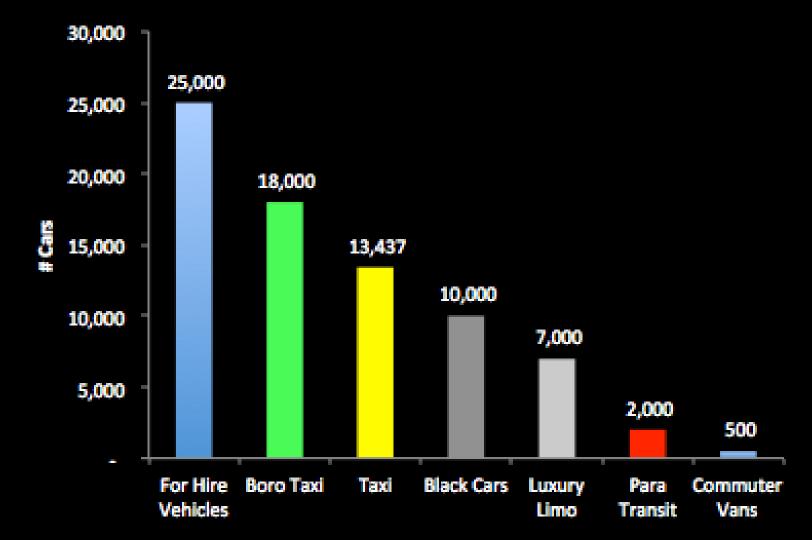


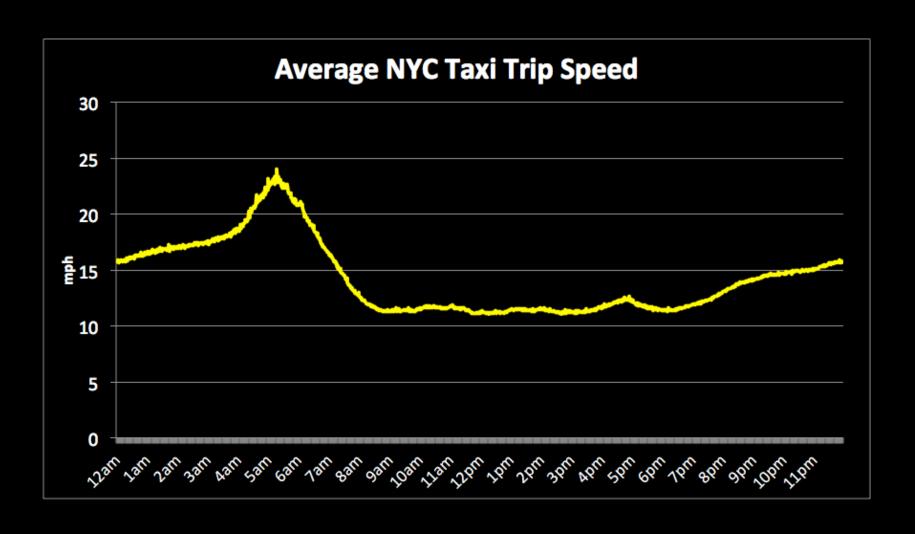
# Sem' Open Data

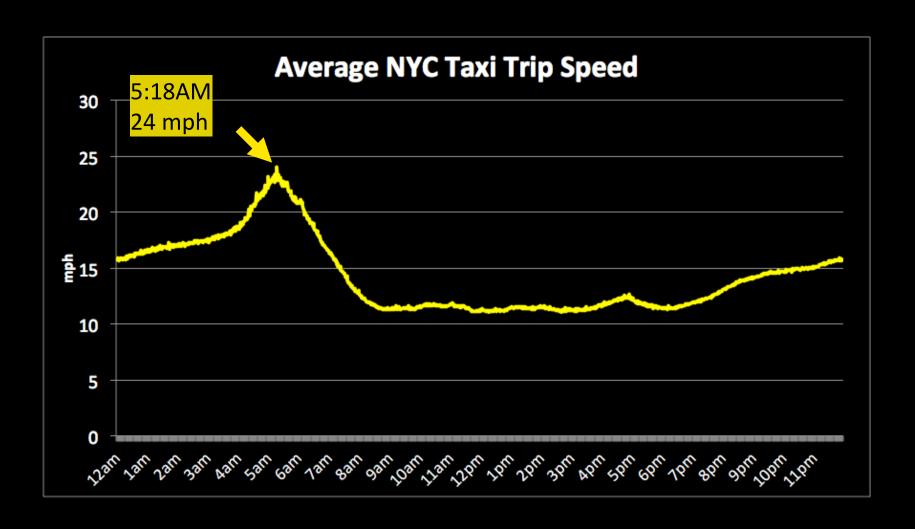
# Uses Valuable Talent to Make "Public" Data Public

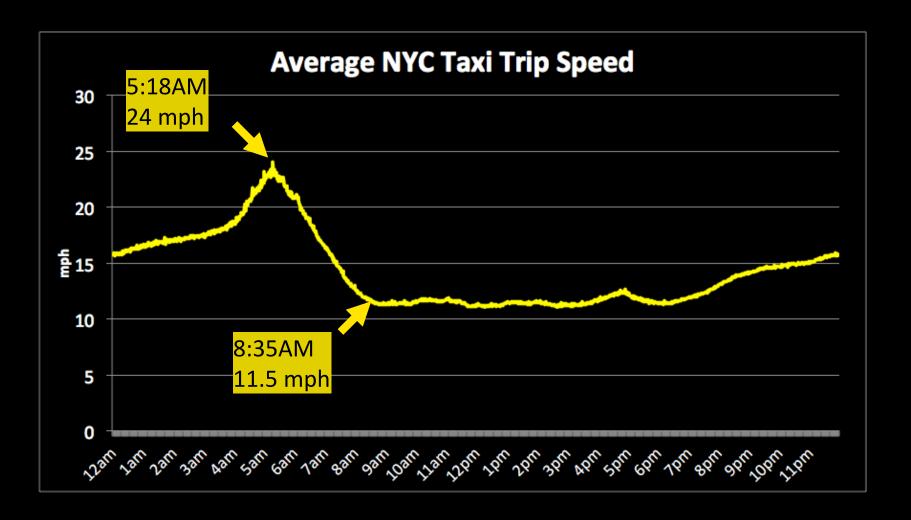
# "Because the files were in PDF format we couldn't sort or analyze the data"

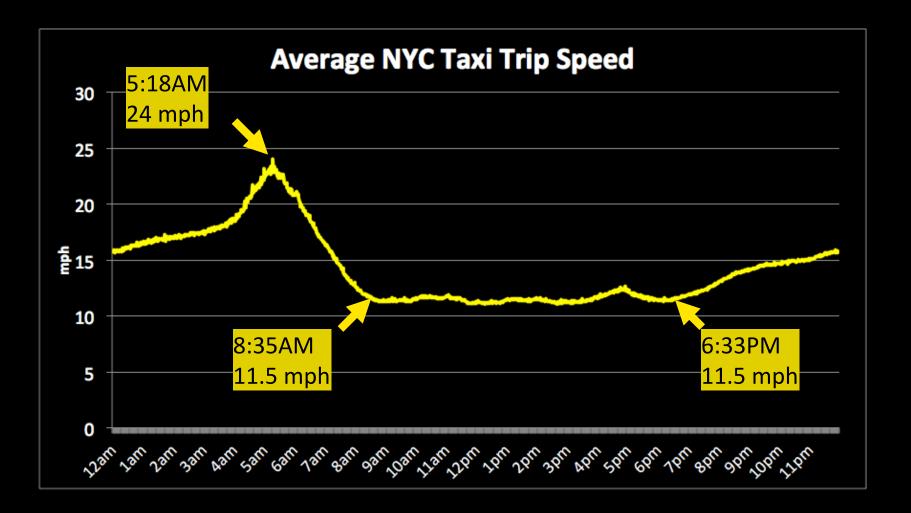
Verizon FIOS Implementation Final Audit Report NYC to Verizon, Jun 18, 2015













Meera Joshi Commissioner

Christopher Wilson Deputy Commissioner/General Counsel Office of Legal Affairs 33 Beaver Street, 22<sup>nd</sup> Floor New York, NY 10004

+1 212 676 1102 fax

#### FOIL REQUEST FORM

Taxi and Limousine Commission
Office of Legal Affairs

n

FROM: (Please print your name, address, telephone # and email address)

33 Beaver Street, 22<sup>nd</sup> Floor New York, New York 10004 Attn: Records Access Officer

Our email address: FOIL@tlc.nyc.gov Your email address: \_\_\_\_

I request the following record(s) under the Freedom of Information Law:

Please reasonably describe the record(s) you are requesting to allow us to identify any responsive document(s).



Meera Joshi Commissioner

Christopher Wilson Deputy Commissioner/General Counsel Office of Legal Affairs 33 Beaver Street, 22<sup>nd</sup> Floor New York, NY 10004

+1 212 676 1102 fax

#### FOIL REQUEST FORM

Taxi and Limousine Commission Office of Legal Affairs

FROM: (Please print your name, address, telephone # and email address)

33 Beaver Street, 22nd Floor New York, New York 10004 Attn: Records Access Officer

Our email address: FOIL@tlc.nyc.gov Your email address:

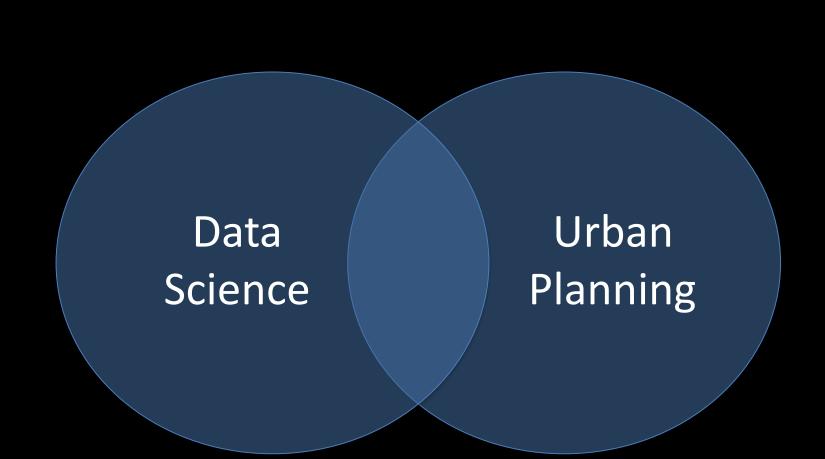
I request the following record(s) under the Freedom of Information Law:

Please reasonably describe the record(s) you are requesting to allow us to identify any responsive document(s).



# Sem' Open Data

Uses Valuable Agency Time to Reply to FOIL Requests



## I Quant NY

MAILING LIST

R

**ARCHIVE** 

Quantitative Analysis of NYC Open Data: Every data set that the city releases tells a story. This blog is all about telling those stories, one data set at a time.

About Me About You Interviews Press Topics Subscribe

SEPTEMBER 5, 2014

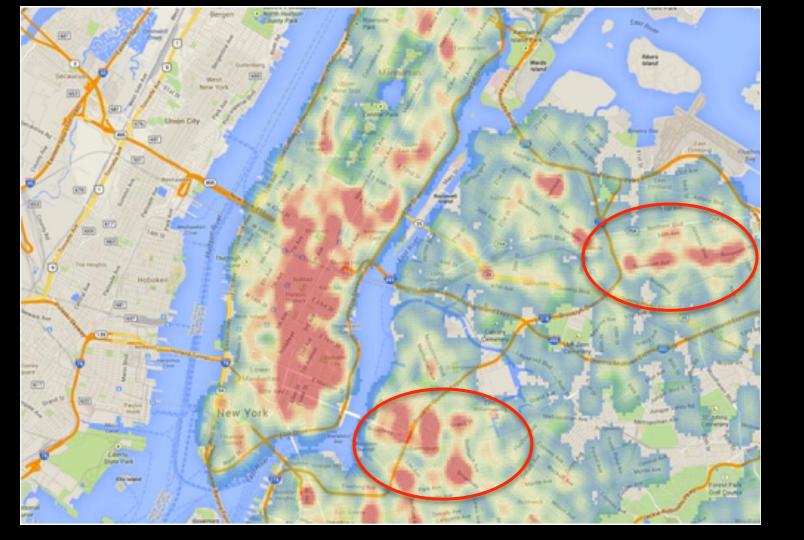
## How Memorizing "\$19.05" Can Help You Outsmart the MTA

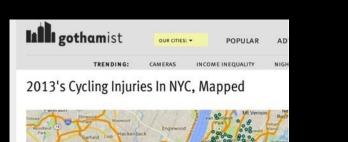
We've all been there. The train is coming into the station, and you grab your MetroCard and quickly try and swipe it at a turnstile.

"Please Swipe Again". "Please Swipe Again". "Insufficient Fare".

The last two words are killer. You think to yourself "I swear I had a balance on this card". You go and check the card out and you see you have "\$2.45". Yes, you need \$2.50 to ride the subway, and you have \$2.45 on your MetroCard. Sure enough you miss that train all because of that nickel.

How did you end up in that situation any way? It turns out the MTA has designed it that way. Imagine how many tourists come to NYC and leave with balances that never get used. Imagine how many people lose metro cards with those balances that never get used. And even if it gets used on a later refill, the MTA gets to collect the cash earlier this way! Win win for them, right?











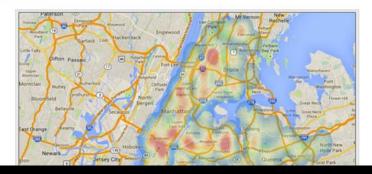
### STREETSBLOG NYC

Transit | Bicycling | Walking | Public Space | Vision Zero | DMV Reform

Thursday, February 27, 2014

D 1 Comme

Heat Maps Show Where Traffic Takes the Most Lives in NYC by Stephen Miller





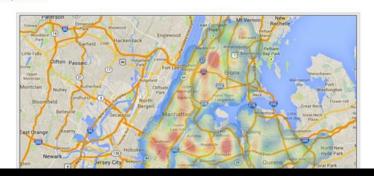


Bicycling Walking Public Space Vision Zero DMV Reform

Thursday, February 27, 2014

D 1 Commi

Heat Maps Show Where Traffic Takes the Most Lives in NYC by Stephen Miller







**PHOTOS** 

WORK HOUSING WEATHER

MAPS

### **Mapping New York's Traffic** Crashes

Twenty-three percent of the city's 2013 car-related fatalities occurred in just 5 percent of its neighborhoods.

SARAH GOODYEAR | 💆 @buttermilk1 | Feb 27, 2014 | 🟴 4 Comments











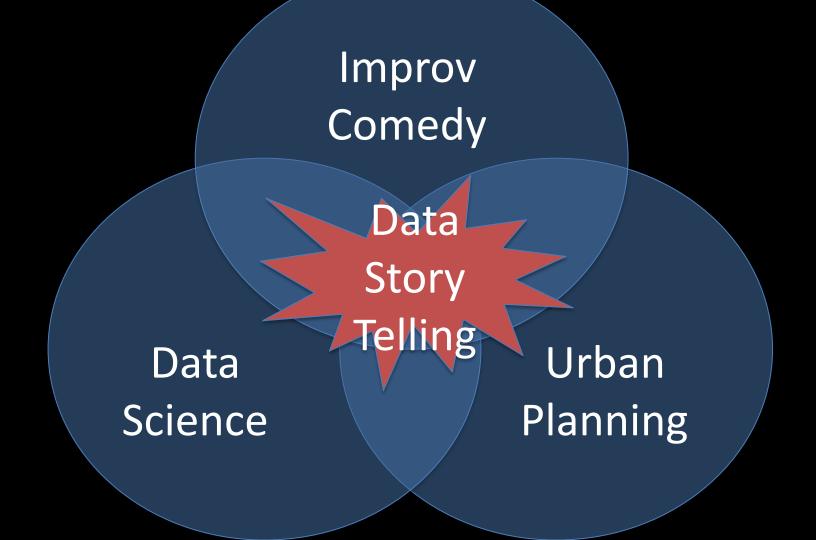


The New York Police Department doesn't make it easy to crunch the city's traffic fatality and injury numbers, releasing the data in a PDF format that's

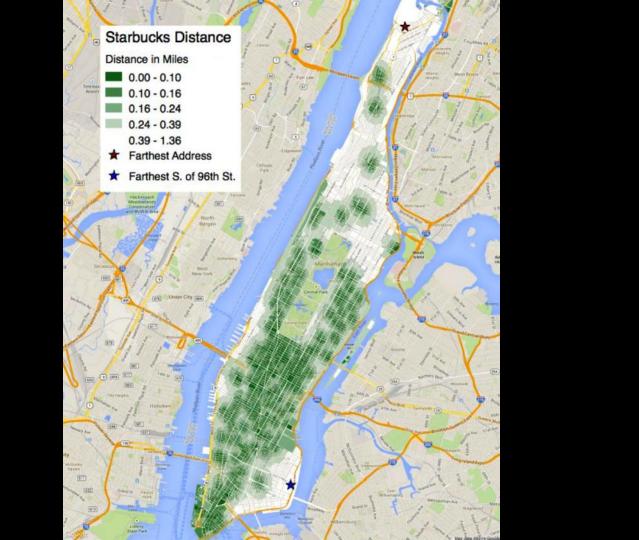
# Urban Data Science Planning

Improv Comedy

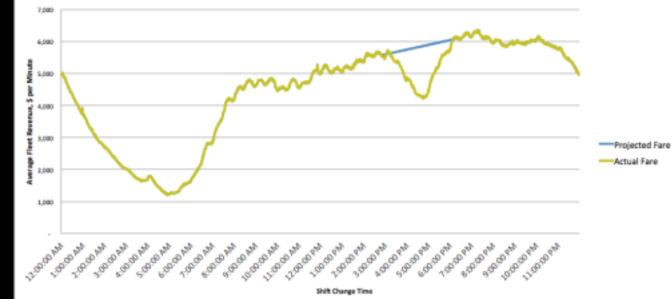
Data Science Urban Planning



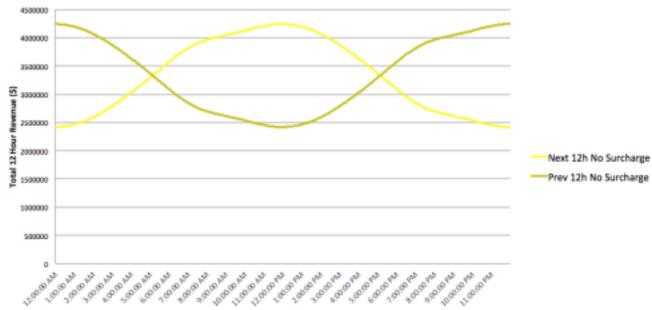
# Connect With People's Experiences



### NYC Taxi Fare Revenue Per Minute, 2013

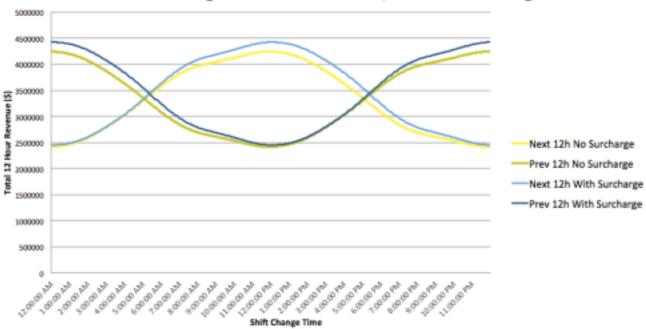


### NYC Taxi Fare Rolling 12 Hour Revenues

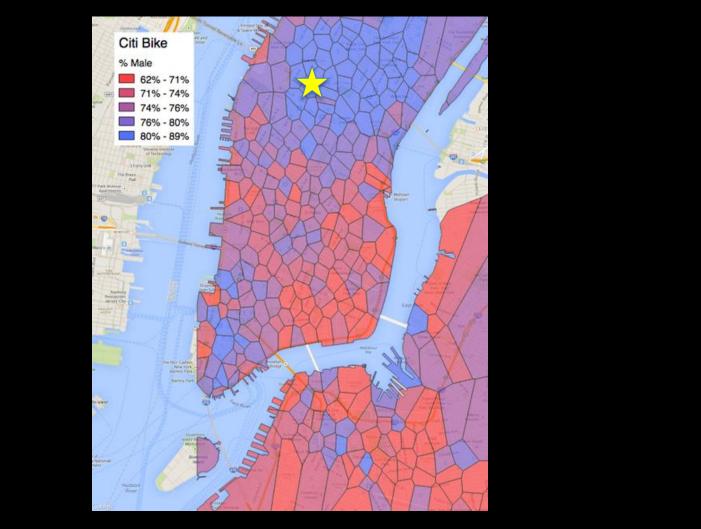


Shift Change Time

### NYC Taxi Fare Rolling 12 Hour Revenues w/ Reverse Surcharge



## Focus on A Single Idea



## Keep it Simple

 $\int \vec{E} d\vec{\ell} = -\iint \frac{\partial \vec{B}}{\partial t} \cdot d\vec{S} \quad \vec{E} = 4 \frac{\rho_1 \rho_2}{r^2} \vec{V} = \iint \vec{D} d\vec{S} = AD \left( \frac{E_t}{E_0} \right)_{\parallel} = \frac{2 \cos \vec{v}_1 \cos \vec{v}_2}{\cos (\vec{v}_1 - \vec{v}_2) \sin (\vec{v}_1 + \vec{v}_2)}$  $= \frac{F_e}{9_0} = k \frac{Q}{r} \int \overrightarrow{B} d\overrightarrow{\ell} = \mu \int \overrightarrow{J} d\overrightarrow{S} \int \overrightarrow{f} = \frac{R_0^2 R_0}{(m-1)(R_0 - R_0)} \frac{m_1}{x} + \frac{m_2}{x'} = \frac{m_2 - m_1}{r} \overrightarrow{S} = \frac{1}{m_0} (\overrightarrow{E} \times \overrightarrow{B})$ 

By = VENU Facin (Kong)

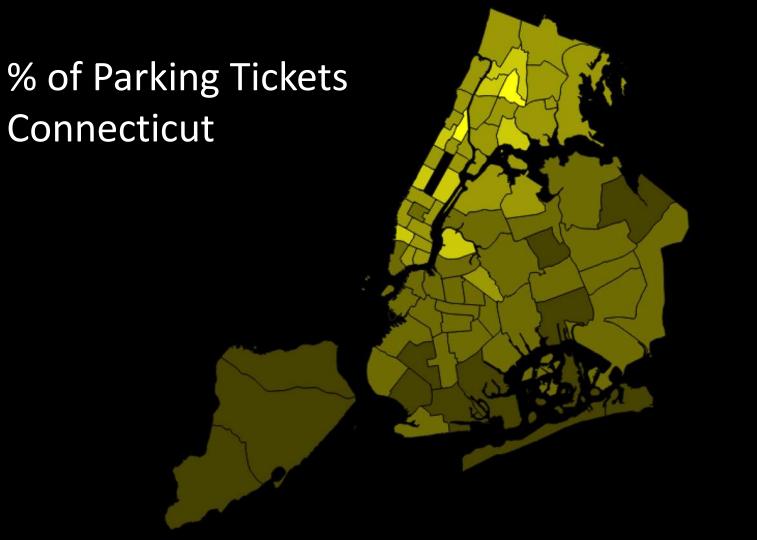








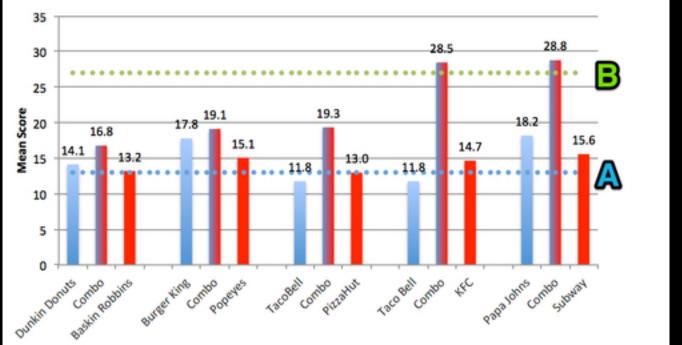






# Explore the Things You Know Best

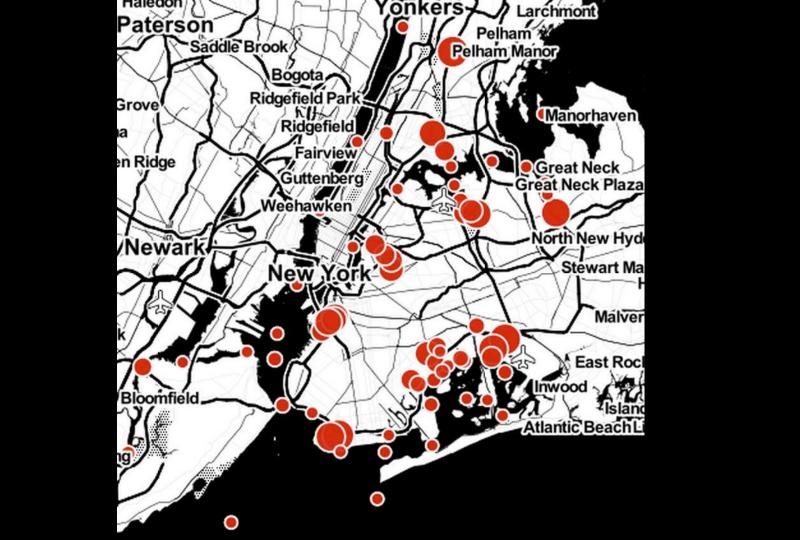
### NYC Restaurant Inspection Scores, 2010-2014



### NYC Restaurant Inspection Scores, 2010-2014



# Figure Out What Exactly You are Trying to Say





## Apartments dumped 200,000 gallons of sewage per day into Coney Island Creek

[+] Enlarge this image

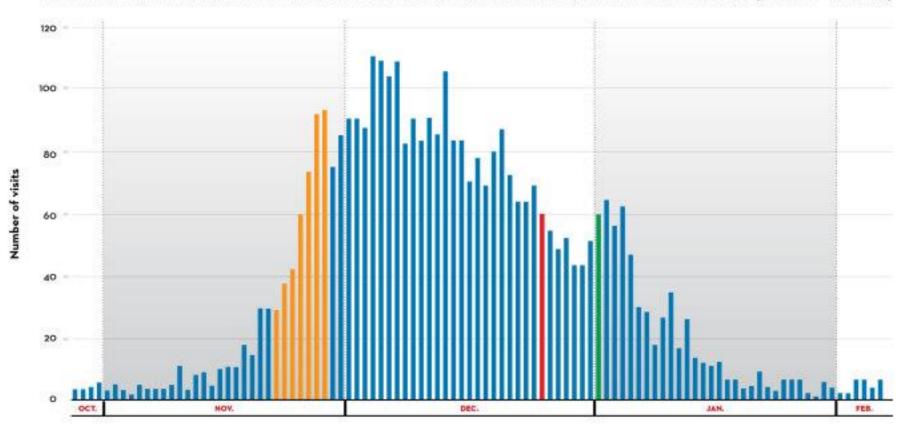


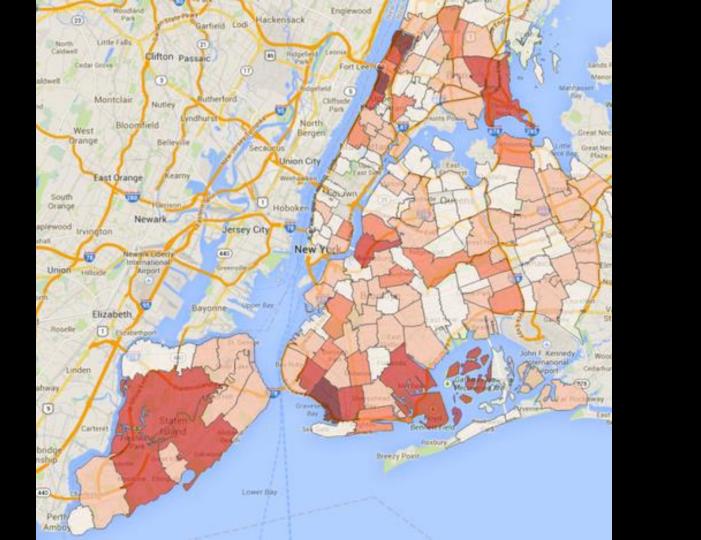
Photo by Georgine Benvenuto

Poop chute: The slurry entered the creek through this storm drain near Shell Road and Shore Parkway.

### Think Outside the Box

### EMERGENCY-ROOM VISITS RELATED TO CHRISTMAS DECORATIONS (2003–2013)





Firework Complaints via 311 25 20 Number of Complains PEr Day

7/1

Day of Year

9/1

10/1

11/1

12/1

1/1

2/1

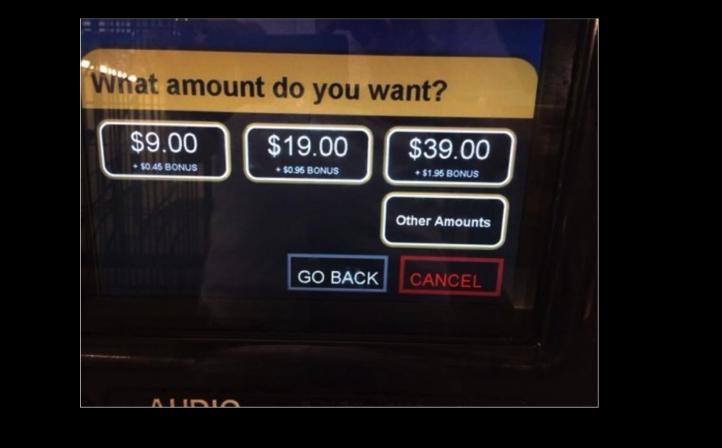
3/1

4/1

5/1

6/1

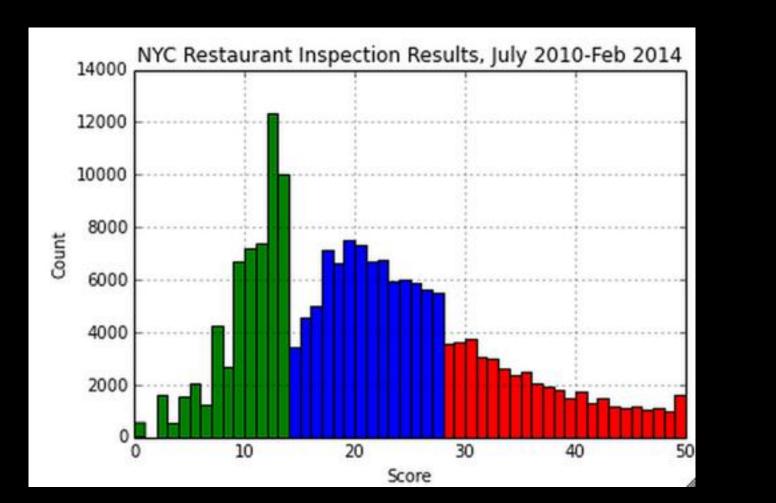
Make an Impact



"These machines do not hold an infinite amount of change and the denominations are suggested to insure there is ample change to accommodate customers who pay with cash. That being said, we will certainly look at this as part of the process involved in rolling out the next scheduled fare increase slated for next year"

-Metropolitan Transit Authority





"Inspectors are not instructed to offer leniency, just to cite what they see. The final score is based on the extent of the violations that the inspector observes."

-NYC Department of Health

## **CMT**

20% of Fare, Surcharge, Tax, Toll

I ♥ NEW YORK RideLinQ Payment Receipt Vehicle # AE837 HACK #: 05414228 Driver# 5179627 MED #: Trip# 2936 12/18/2014 10:00-10:18 Date 11/24/2014 TRIP #: 9:32 AM to 9:49 AM STAND. CITY RATE **RATE #: 1** Distance 4.1 miles MILES R1: Fare: FARE R1: Rate 1: \$15.50 EXTRAS: Tolls: ST. SUR: \$0.00 TIPS: Surcharges \$0.00 TOTAL: Extra CARDNUMBER: \$0.00 AUTHOR .: Tax \$0.50 Tip \$3.20 Contact TLC Dial 3-1-1 Total \$19.20

### Verifone

20% of Fare, Surcharge

AE417

3301

4.15

\$ 16.00

\$ 0.00

\$ 0.50

\$ 3.20

\$ 19.70

76542C

"We appreciate the work that went into his analysis, and we're giving it a thorough read"

-NYC Taxi and Limousine Commission





The 5th Pct. had a 115% increase in bicycle collisions this year Police Officers will be enforcing violations by bicyclists.#bike safety

3:23 PM - 21 Nov 2014

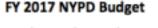


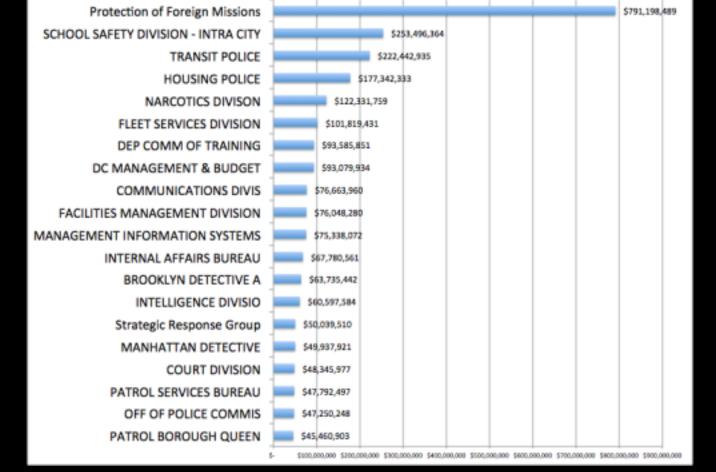




u n

-NYPD





--NYC Mayor's Offce

and on Open Data"

"We are working on fixing this misprint in

the publicly shared Supporting Schedule







"While DOT has not received any complaints

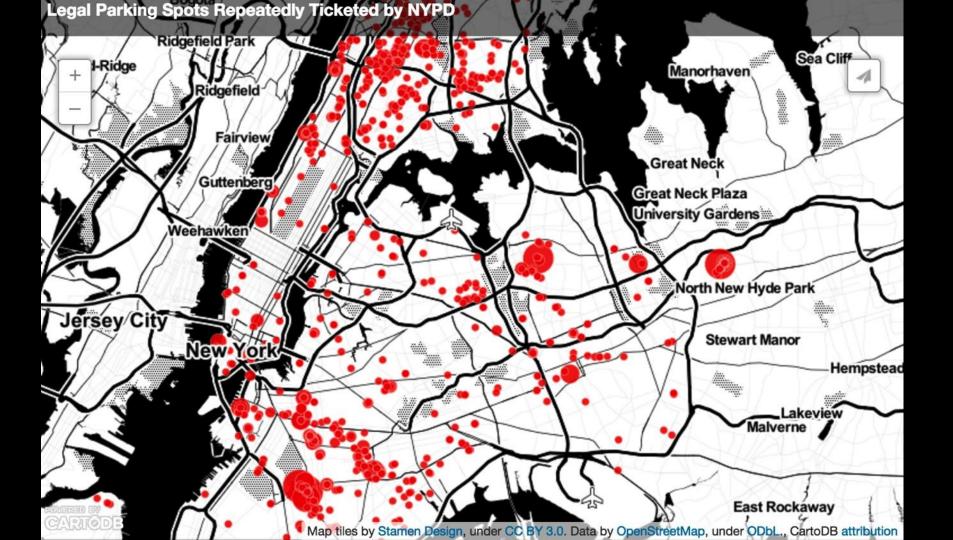
about this location, we will review the roadway

markings and make any appropriate alterations."

-NYC Department of Transportation







"Mr. Wellington's analysis identified errors the department made in issuing parking summonses. It appears to be a misunderstanding by officers on patrol of a recent, abstruse change in the parking rules. We appreciate Mr. Wellington bringing this anomaly to our attention."

-NYPD

# Data Storytelling

- Connect with People
- Try to Convey One Idea
- Keep it Simple
- Explore the Things You Know Best
- Figure Out What Exactly You are Trying to Say
- Think Outside the Box







40.8571285

40.7216199

40.6875062

40.5584106

AN 723681









Sign Up Sig

-73.8807926

-73.7767096

-73.9146441

-74.1693169

-73 ROOK07K

NYC	0	pen	Data

11 🗮 02/13/2015

12 🔚 02/13/2015

13 🔚 02/13/2015

14 🗮 02/13/2015

15 🔚 02/13/2015

16 🗮 02/13/2015

17 🗮 02/13/2015

18 := 02/13/2015

	124	UT	
Data	sets	Availal	ole

	Details of Motor venicle Collisions in New York City provided by the Police Department (NYPD).						
	DATE <b>⊕</b> ∷	TIME <b>⊕</b> ∷	BOROUGH <b>⊕</b> ∷	ZIP CODE <b> </b>	LATITUDE <b> </b>	LONGITUDE	
	*						
1	02/13/2015	0:01					
2	02/13/2015	21:45	MANHATTAN	10002	40.7156221	-73.9942752	
3	02/13/2015	21:45	MANHATTAN	10001	40.7475349	-73.9883068	
4	02/13/2015	21:45	BRONX	10462	40.8335582	-73.8577325	
5	02/13/2015	21:44	MANHATTAN	10017	40.7487997	-73.969846	
6	02/13/2015	21:40	STATEN ISLAND	10304	40.6172954	-74.0804791	
7	02/13/2015	21:39	BRONX	10462	40.8319276	-73.85104	
8 :	02/13/2015	21:33	MANHATTAN	10016	40.7464777	-73.9838968	
9 :	02/13/2015	21:25	BROOKLYN	11237	40.705679	-73.9260903	
10	02/13/2015	21:20	MANHATTAN	10036	40.7615555	-73.9940691	

10458

11432

11221

10312

11278

21:20

21:15

21:15

21:15

21:15

21:11

21:07

21.05

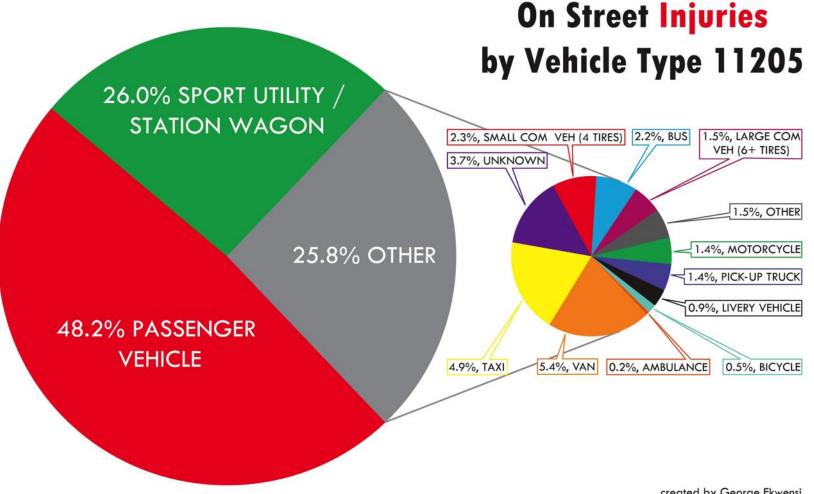
**BRONX** 

**QUEENS** 

OLIFFNIC

**BROOKLYN** 

STATEN ISLAND



created by George Ekwensi SOURCE: NYPD, NYC OPEN DATA







NYPD

NYPD









Q Find in this Dataset

Sign Up S

29952438

29947408

### 311 Service Requests from 2010 to Present

All 311 Service Requests from 2010 to present. This information is automatically updated daily.

All 31	1 Service Requests from 2010 to present. This inform	ation is automatically updated daily.	★ Manage  More Views	<b>T</b> Filter	✓ Visualize Export 19 Discuss () Embed	
	Unique Key <b> </b>	Created Date	Closed Date	Agency	Filter	
1	29952579	02/16/2015 02:12:30 AM		NYPD	Conditional Formatting	
2 :≣		02/16/2015 02:10:15 AM	02/16/2015 03:34:36 AM	NYPD	Sort & Roll-Up	
3 ≣	29947340	02/16/2015 02:10:04 AM		DOT	Filter	
4	29952664	02/16/2015 02:07:52 AM		TLC	Filter this dataset based on contents.	
5 📜	29953428	02/16/2015 02:07:43 AM		NYPD		
6 ≣	29953346	02/16/2015 02:07:08 AM	02/16/2015 03:12:59 AM	NYPD	Unique Key v is v	
7 📜	29947201	02/16/2015 02:01:26 AM	02/16/2015 02:45:01 AM	NYPD		
8 ≣	29947469	02/16/2015 01:58:36 AM		NYPD	Ot-	
9 <b>≣</b>	29948047	02/16/2015 01:55:34 AM		NYPD	+ Add a New Filter Condition	
10 등	29947486	02/16/2015 01:53:38 AM	02/16/2015 02:18:56 AM	NYPD		
11 등		02/16/2015 01:47:38 AM		NYPD	Never created a filter before? Watch a short tu video here.	
12 🧮		02/16/2015 01:47:20 AM		NYPD		
13 📜		02/16/2015 01:47:16 AM		NYPD		
14		02/16/2015 01:39:41 AM		NYPD		
15	29952644	02/16/2015 01:37:16 AM	02/16/2015 02:54:21 AM	NYPD		

02/16/2015 01:37:13 AM

02/16/2015 01:32:35 AM

AMOUNT OF EGRAFFITI COMP-LAINTS (BY ZIP CODE) 300 + complaints 200-300 comprounts 120-199 complaints H-119 complaints 36-75 componints 0-35 complaints SAY. Charle. abby Shosid

## Anyone can be a Data Storyteller.

iquant.nyc

@iquantny