



Data Analytics for Managers

Instructor: Richard Dunks

27 July 2016

Datapolitan

Data Solutions for the Modern Metropolis

Goals for the Course

- Discuss the process of data-driven decision making as it relates to city government
- Introduce useful terminology around data and the data analytics process
- Explore examples of good analytics efforts in US cities and get some hands-on experience analyzing data
- Explore the value of data, especially open data in the analytics process

Key Takeaways for the Course

- You will better understand the value of data, particularly government open data, in the decision-making process
- You will better understand the analytics process
- You will better understand how to build a data-driven culture
- You will be more familiar with examples of analytics in NYC government and other cities around the US

Goals for this Morning

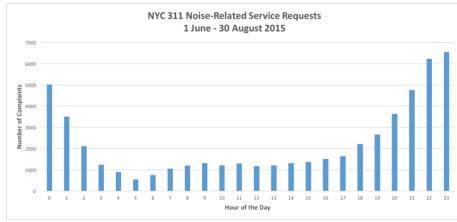
- Discuss the concept of “data-driven”
- Discuss types of analysis in city government
- Discuss the benefits and concerns around data analytics in operational decision making
- Apply an understanding of the analytic process to a New York City-specific problem
- Discuss the features of a data-driven culture

Why Do We Collect Data?

- Accountability
- Transparency
- “Can’t manage what you can’t measure”

The Value of Data

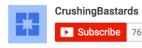
- Data describes phenomena of interest
- Can describe the phenomena directly or indirectly



ARE ALL DATA POINTS CREATED EQUAL?



Girls Crash into Lake following Bad GPS directions



CrushingBastards

Subscribe

767

192,977

224

22

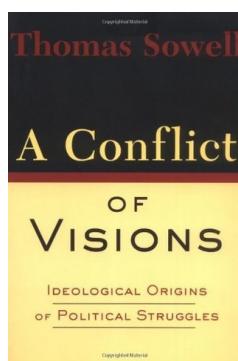
Uploaded on Jun 15, 2011

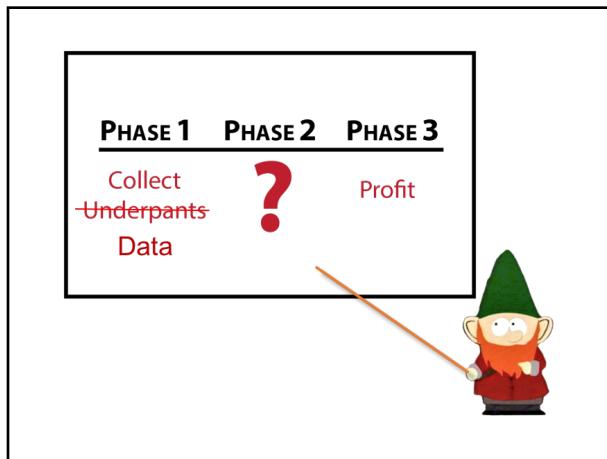
Three young women escaped a sinking SUV after a direction from a rental car GPS unit sent them down a boat launch and into the Mercer Slough early Wednesday.

<https://www.youtube.com/watch?v=a20H2uz3p8>

Facts do not "speak for themselves." They speak for or against competing theories. Facts divorced from theory or visions are mere isolated curiosities.

— Thomas Sowell,
A Conflict of Visions







WHAT DOES “DATA-DRIVEN” MEAN IN THE CONTEXT OF CITY GOVERNMENT?

You feelin' lucky, bud?
Go ahead, say "data-driven"
one more time and let's
see what happens.

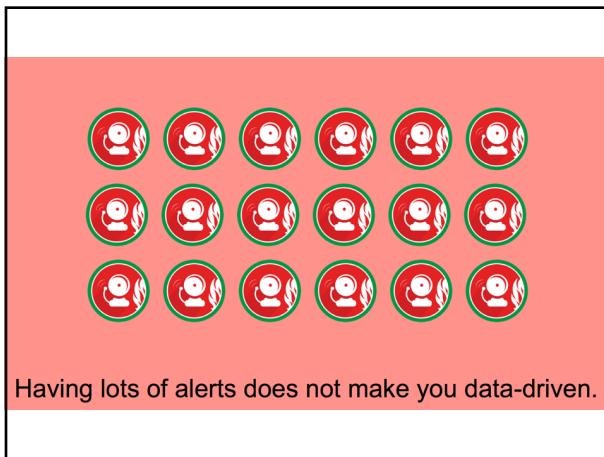


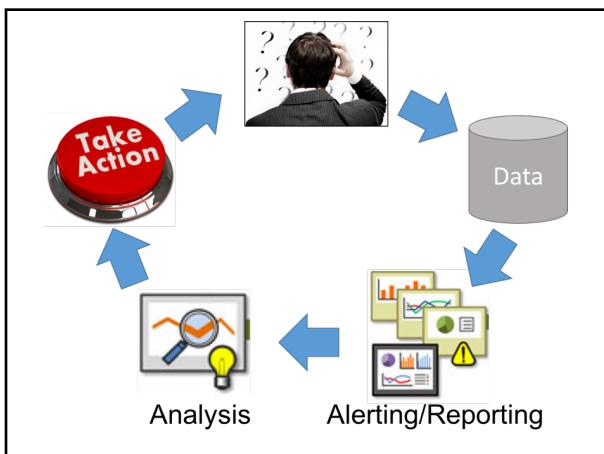


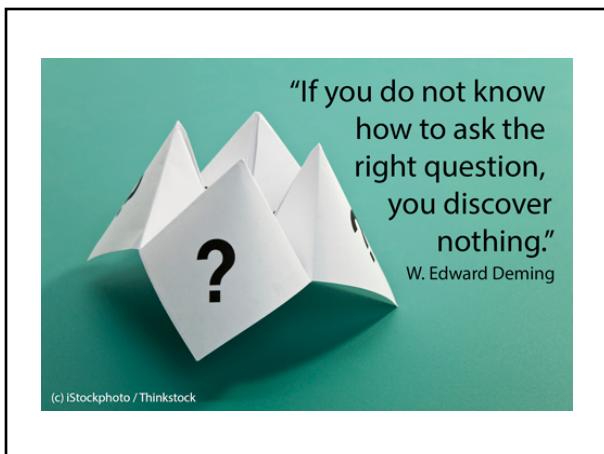
Having lots of reports does not make you data-driven.



Having lots of dashboards does not make you data-driven.









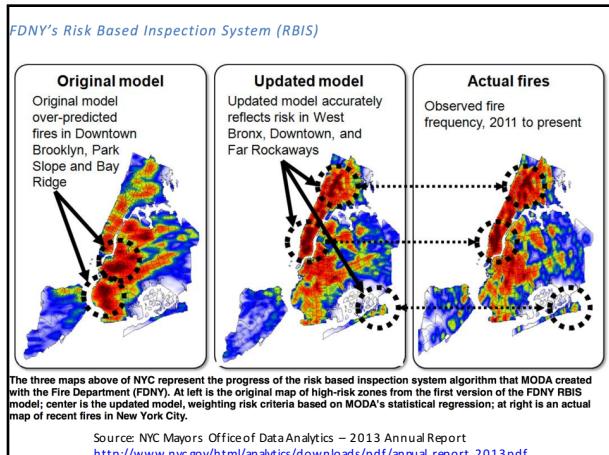
"Errors using inadequate data are much less than those using no data at all."
Charles Babbage

(c) Digital Vision / Thinkstock

TYPES OF ANALYSIS

1. Quantifying Needs

- How much of X do I need?
 - Analyzing inputs (resources, people, etc.)
- How much does my need change given a different set of conditions?
 - What are the conditions that influence X?
- Important Considerations:
 - How does X play into my organization's mission and goals?
 - What's the most meaningful way of quantifying X?



2. Operational Analysis

- What is my organization doing?
 - Assessment
- How might my organization do things better?
- Important Considerations:
 - What are your organization's mission and goals?
 - How do your employees do their work?
 - What's the best way to measure this work?

64° Sign In | Subscribe

The Charlotte Observer

BUSINESS AUGUST 28, 2015

Cincinnati racks up more than \$130,000 in late fees to Duke Energy

To address the issue, workers in Cincinnati's new Office of Performance and Data Analytics identified the impact the fees were having on the city and spent three days this summer with city department heads to come up with a solution, the station reported.

City leaders spent time in what's known as the Innovation Lab, where they figured out a "new way" to pay bills on time and avoid such fees. It wasn't immediately clear what their solution was.

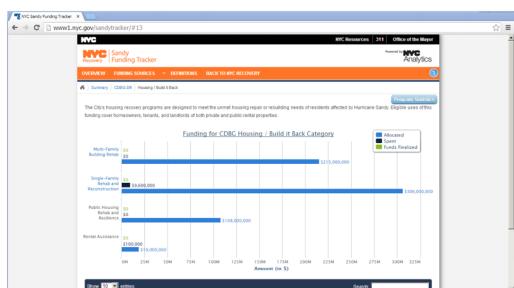
Source: <http://www.charlotteobserver.com/news/business/article32617293.html>

3. Performance Metrics

- How is my organization doing?
 - Monitoring and evaluation
- How do we make this data visible to the people who need it?
- Important Considerations:
 - What is most important to measure (think mission and goals)?
 - How do we best measure performance?

Tracking Hurricane Sandy Relief Funds

Tracking information on Sandy recovery funds, built by NYC analytics, is available at www1.nyc.gov/sandytracker



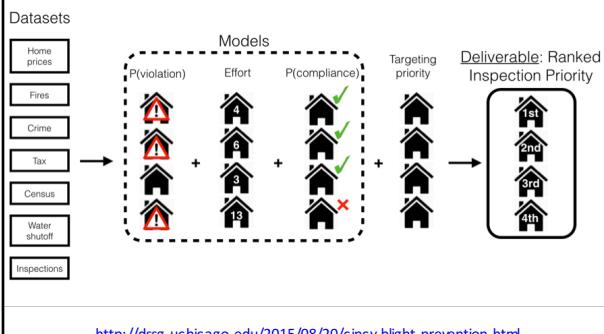


Source: <http://www.slideshare.net/LeanAnalytics/startup-metrics-toronto-march-19>

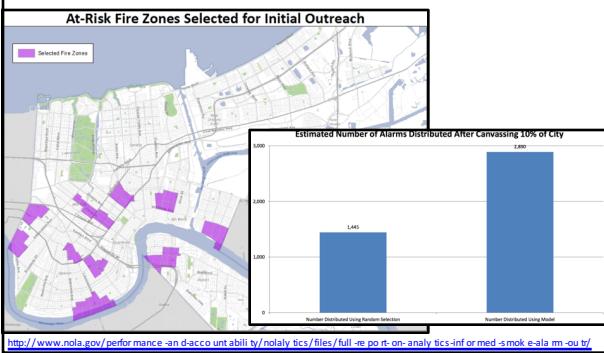
4. Prioritization

- How do I meet optimal outcomes with limited resources?
 - Optimizing allocation
- Important considerations:
 - Minimize disruption
 - Work within current workflow
 - Support existing business practices

Cincinnati Blight Prevention Model



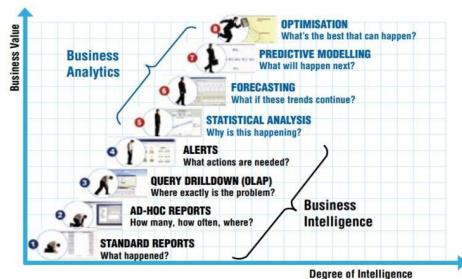
New Orleans Smoke Alarm Targeted Outreach



5. Data Sharing/Empowering Stakeholders

- How could others benefit from my data?
- What other data can I use?
- Important Considerations:
 - Machine-readable formats
 - Make your data “fit” with other data sources
 - Unique IDs
 - Indexes
 - Key values

Levels of Analytics



Source: https://www.sas.com/news/sascom/analytics_levels.pdf

Being data-driven doesn't mean



blindly following data.

Augment decision makers with objective, trustworthy, and relevant data.

Intel chips could have powered first iPhone, CEO Otellini says

By AppleInsider Staff

Thursday, May 16, 2013, 02:42 pm PT (05:42 pm ET)

As Paul Otellini relinquishes the reins of chipmaker Intel, the outgoing chief executive reveals in an in-depth profile that he ultimately passed on a contract to build the silicon which powered Apple's original iPhone, a regretful decision given the handset's wild success.



Retiring Intel CEO Paul Otellini.

<http://appleinsider.com/articles/13/05/16/intl-chips-could-have-powered-first-iphone-ceo-otellini-says>

4 TYPES OF CONCERNS TO BE MINDFUL OF

1. Technical

- Having the right tools
- Having the people who can use them
- Making everything work together
- *Potential trap: having a solution in search of a problem*

2. Legal

- Laws
- Regulations
- Practices
- *Potential trap: not doing something because of mistaken assumptions*

The New York Times

HEALTH

Hipaa's Use as Code of Silence Often Misinterprets the Law

JULY 17, 2015

Paula Span
THE NEW OLD AGE

How do people use, misuse or abuse Hipaa, the federal regulations protecting patients' confidential health information? Let us count the ways:

Last month, in a continuing care retirement community in Ithaca, N.Y., Helen Wyvill, 72, noticed that a friend hadn't shown up for their regular swim. She wasn't in her apartment, either.



http://www.nytimes.com/2015/07/21/health/hipaa-use-as-code-of-silenceoften-misinterprets-the-law.html?_r=0

3. Cultural

- "We've always done it this way"
- "I'm not sure I understand how this works"
- *Potential trap: being afraid of rocking the boat*

4. Political

- Inter-departmental
- Intra-departmental
- *Potential trap: not putting the necessary effort into something that will pay dividends to your agency and ultimately to the city as a whole*

Political Example – inBloom

- Non-profit company founded in 2011 by Council of Chief State School Officers
 - Supported with funding from the Bill and Melinda Gates Foundation, among others
- Sought to provide an open-source platform for combining data from various education vendor products
- Educators could use data in one consolidated system to improve learning

Political Example – inBloom

- Public concern over the potential use of the data by 3rd parties led states to cancel contracts
- The company began winding down operations in April 2014
- Lesson in how politics must be factored in – inBloom lost in the court of public opinion

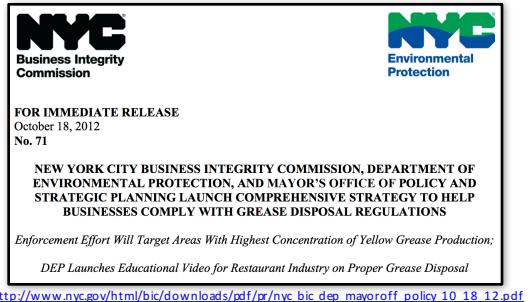
<http://www.businessweek.com/articles/2014-05-01/in-bloom-s-huts-down-in-a-mid-privacy-feas-vers-tudent-data-tracking>

Benefits

- Time, money, lives saved
- Better delivery of services to stakeholders
- More transparency
- More accountability

Benefits

- Time, money, lives saved



I Quant NY

[MAILING LIST](#) [RSS](#) [ARCHIVE](#)

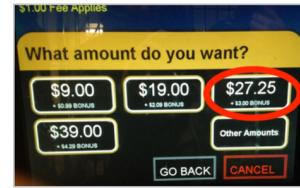
Qualitative Analysis of NYC Open Data: Every data set that the city releases tells a story. This blog is all about telling those stories, one data set at a time.

[About Me](#) [About You](#) [Interviews](#) [Press](#) [Topics](#) [Subscribe](#)

MARCH 24, 2015

I Quant a Victory: MTA Adds New Button for No-Left-Over-Balance MetroCards!

First, they have released a new fare calculator which allows you to determine how to buy or refill a card to get even balances. That is pretty exciting, though odds are that most people, especially tourists, will not be using it day-to-day when they make purchases. Still, it's a step in the right direction as far as transparency goes.



<http://iquantny.tumblr.co/post/11447010120/i-quant-a-victory-mta-adds-new-button-for-no-left-over-balance-metrocards>

I Quant NY

[MAILING LIST](#) [RSS](#) [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](#)

JUNE 2, 2014

Success: How NYC Open Data and Reddit Saved New Yorkers Over \$55,000 a Year

Before Open Data:



After Open Data:

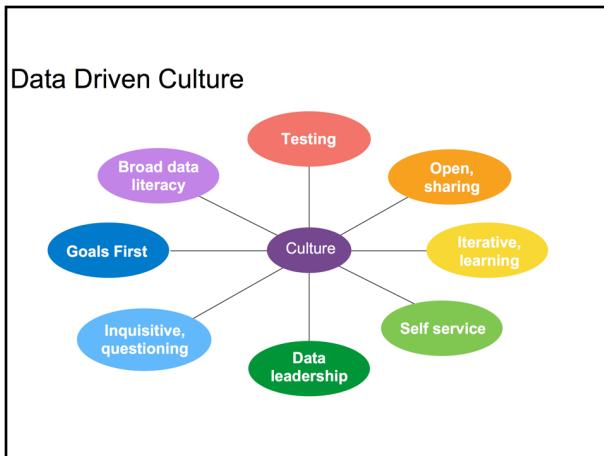
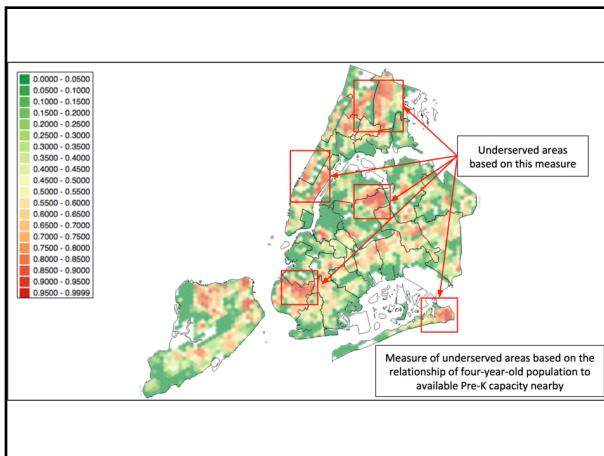
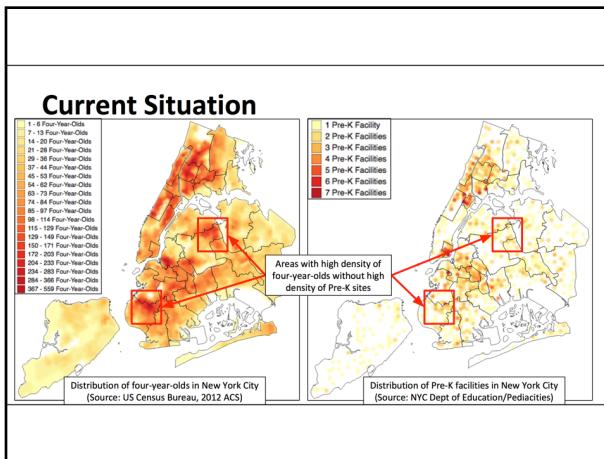
<http://iquantny.tumblr.com/post/87573867759/success-how-nyc-open-data-and-reddit-save-new-yorkers-over-55000-a-year>

WHAT CONCERNS DO YOU HAVE WITH RESPECT TO ANALYTICS IN YOUR JOB?

WHAT ARE SOME OF THE BENEFITS OF GOOD ANALYTICS IN YOUR OFFICE?

Group Exercise – Universal Pre-K

- Define the analytical problem and data needs in one of these key areas
 - Capacity
 - Outreach
 - Enrollment (CBOs/Students)
 - Monitoring/Evaluation
- Situation
 - ~104,000 4-year olds in NYC
 - 58,528 current seats
 - 26,364 in public schools
 - 32,164 in community based organizations
- Goals
 - Increase enrollment by 30,000 for 2014-2015
 - Increase enrollment by 20,000 for 2015-2016



Being data-driven means having...



a self service culture

Business units have necessary data access as well as within-team analytical skills to drive insights, actions, and impact.

“Central Source of Truth”



Invest in data quality

1 / 3

business leaders frequently make decisions with data that they cannot trust

2009. Business Analytics and Optimization for the Intelligent Enterprise. IBM

Are you sure you want to move on to problem solving?

I vote we dispute the accuracy of these budget summary spreadsheets for a few more hours.



your ecards
someecards.com

Being data-driven means having...



a broad data literacy

All decision-makers have appropriate skills to use and interpret data.

Being data-driven means having...



an objective, inquisitive culture

"Do you have data to back that up?" should be a question that no one is afraid to ask and everyone is prepared to answer'—Julie Arsenault.

Being data-driven means having...



strong data leadership

A head of data to evangelize data as strategic asset with budget, team, and influence to drive cultural change.

Change should not just be top-down



but bottom up too

Everyone in org has role and responsibility through "leveling up" their data skills, mutual mentoring, and embedding data into their processes.

Being data-driven means having...



an open, sharing culture

No data hoarding or silos. Bring data together to create rich contexts. Connect the dots.

Culture

Collaborative, inclusive, open, inquisitive

Data leadership

Chief data officer / chief analytics officer

Decision making

Testing mindset, fact-based, anti-HIPPO

Organization

Embedded, federated analytics

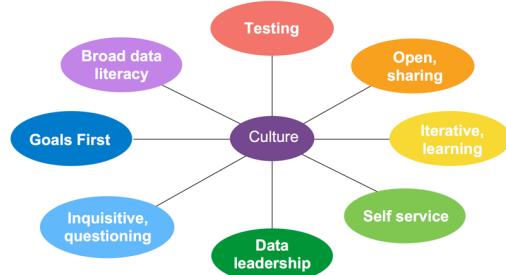
People

Analytics org: composition, skills, training

Data

Data quality, data management

Data Driven Culture



Goals for this Afternoon

- Explore real-world applications of data to operational problems
- Familiarize you with exploratory data analysis and question-driven analysis
- Practice communicating analytical findings
- Become familiar with sources of data available online

6 ANALYTICAL STEPS

1. Problem Formulation

- What question or need am I trying to answer?
- What's my organization's mission and goals?
- How can I best apply data to this task?

2. Data Gathering/Preliminary Analysis

- What data do I think I'm going to need?
- What condition is it in?
- Does it tell me what I need?
- What other data might I need?
- How much work do I need to put into the data?

3. Data Cleaning/Data Munging

- Make the data usable and compatible
- Takes up the most amount of time
- May require more sophisticated tools depending on the state and size of the data

≡ SECTIONS ⌂ HOME 🔍 SEARCH

The New York Times

TECHNOLOGY

For Big-Data Scientists, ‘Janitor Work’ Is Key Hurdle to Insights

By STEVE Lohr AUG. 27, 2014



Monica Rogati, Jawbone's vice president for data science, with Brian Wilt, a senior data scientist.
Peter DaSilva for The New York Times

<http://www.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-to-insights-is-janitor-work.html>

4. Hypothesis Testing

- Am I getting the results I'd hoped for?
 - What other questions come up?

5. Verification

- Do my results make sense?
 - Did I make a simple mistake?
 - Check twice and you'll sleep easier

5. Verification – London Whale

- \$6.2 billion lost by JP Morgan Chase & Co

A screenshot of a news article from Business Insider. The header includes the site's logo, navigation links for Tech, Money & Markets, Briefing, Ideas, Executive Life, and Video. Below the header is a section titled "MONEY & MARKETS". The main headline reads "How The London Whale Debacle Is Partly The Result Of An Error Using Excel". Below the headline is author information ("BY LINETTE LOPEZ | FEB 12 2013, 8:04 AM") and social sharing options. The central content area features a large screenshot of a Microsoft Excel spreadsheet. A red box highlights a specific cell containing the formula '=SUM(A1:A100)'. A tooltip or callout box is overlaid on this cell, pointing to the formula bar with the text "The formula was copied across the entire column". The bottom of the screenshot shows the Windows taskbar with icons for File Explorer, Task View, and other open applications.

<http://www.businessinsider.com.au/excel-partly-to-blame-for-trading-loss-2013-2>

5. Verification – London Whale

- Caused largely by Excel mistakes:
 - Manual data errors
 - Manual copy and paste
 - Simple formula error that hid volatility
 - Fined over \$1 billion for poor internal oversight of trading activities

MarketWatch    

NEWS VIEWER **MARKETS** **INVESTING** **TRADING DECK** **PERSONAL FINANCE** **RETIREMENT** **ECONOMY** **REAL ESTATE**

Home · Personal Finance

88% of spreadsheets have errors

Published: Apr 20, 2013 7:39 a.m. ET

    139

Research: Bad math rampant in family budgets and Harvard studies

 By **JEREMY OLSHAN** EDITOR  

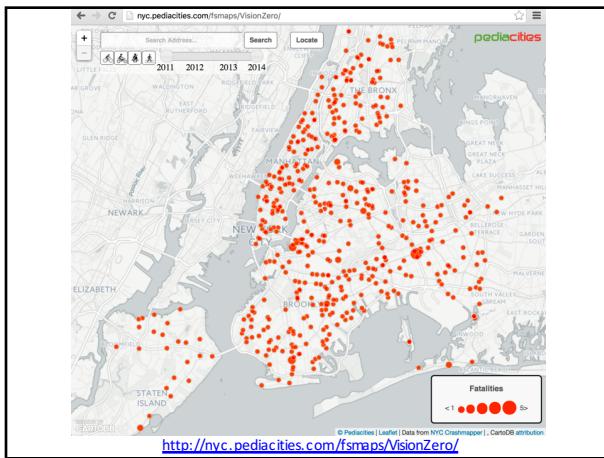
Microsoft Excel makes it easy for anyone to do the kind of number crunching once reserved for accountants and statisticians. But the world's best-selling spreadsheet software has also contributed to the proliferation of bad math.

Close to 90% of spreadsheet documents contain errors, a 2008 analysis of multiple studies suggests. "Spreadsheets, even after careful development, contain errors in 1% or more of all formula cells," writes [Ray Panko](#), a professor of IT management at the University of Hawaii and an authority on bad spreadsheet practices. "In large spreadsheets with thousands of formulas, there will be dozens of undetected errors."

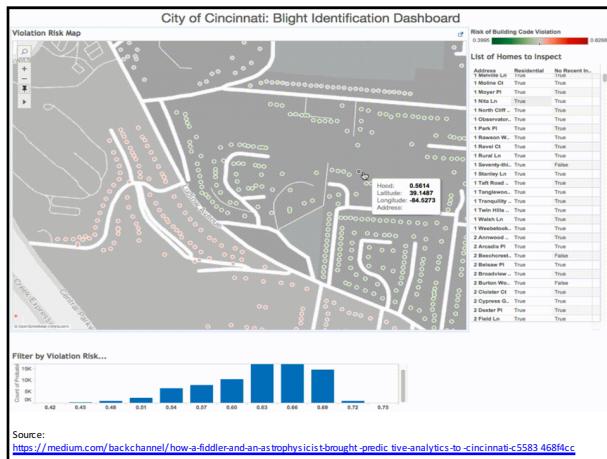
<http://www.marketwatch.com/story/88-of-spreadsheets-have-errors-2013-04-17>

6. Visualization

- “A picture is worth a thousand words”
- Communicate results clearly and concisely
- Help to better understand your data
- The eyes have a much higher bandwidth into the brain



DATE	TIME	BOROUGH	ZIP CODE	LATITUDE	LONGITUDE	LOCATION	ON STREET NAME	CROSS STREET NAME
7/3/12	16:30	BROOKLYN	11216	40.685846	-73.947310	(40.685846, -73.947310)	MARCY AVENUE	MONROE STREET
7/3/12	20:55	QUEENS	11369	40.764014	-73.881215	(40.764014, -73.881215)	ASTORIA BOULEVARD	88 STREET
7/3/12	3:30	BROOKLYN	11225	40.662763	-73.957216	(40.662763, -73.957216)	BEDFORD AVENUE	STERLING STREET
7/3/12	21:54	BROOKLYN	11207	40.751042	-73.903015	(40.751042, -73.903015)	BUSHWICK AVENUE	CONWAY STREET
7/3/12	23:17	BROOKLYN	11223	40.725116	-73.971747	(40.725116, -73.971747)	CLYDE AVENUE	ELIZABETH STREET
7/3/12	0:57	BRONX	10461	40.843075	-73.848547	(40.843075, -73.848547)	OVERING STREET	EAST TRENTON AVENUE
7/3/12	23:18	BRONX	10467	40.878645	-73.871598	(40.878645, -73.871598)	WEBSTER AVENUE	EAST GUN HILL ROAD
7/4/12	21:20	MANHATTAN	10225	40.795367	-73.970369	(40.795367, -73.970369)	BROADWAY	WEST 20 STREET
7/4/12	1:21	BRONX	10456	40.8387216	-73.913770	(40.8387216, -73.913770)	GRAND CONCOURSE	EAST 170 STREET
7/4/12	23:17	BROOKLYN	11221	40.751042	-73.903015	(40.751042, -73.903015)	WEBSTER AVENUE	EAST GUN HILL ROAD
7/4/12	23:22	BROOKLYN	11223	40.650954	-73.926323	(40.650954, -73.926323)	WEINSTEIN AVENUE	AVENUE P
7/4/12	22:51	QUEENS	11104	40.739063	-73.926157	(40.739063, -73.926157)	GREENPOINT AVENUE	39 PLACE
7/2/12	20:08	QUEENS	11434	40.685278	-73.777239	(40.685278, -73.777239)	MERRICK BOULEVARD	VICTORIA ROAD
7/2/12	20:48	BROOKLYN	11212	40.663105	-73.905179	(40.663105, -73.905179)	MOTHER GASTON BOULEVARD	LIVONA AVENUE
7/2/12	3:20	QUEENS	11434	40.685278	-73.777239	(40.685278, -73.777239)	MERRICK EXPRESSWAY	QUEENSBURY EXPRESSWAY
7/2/12	22:15	STATEN ISLAND	10314	40.621486	-73.945646	(40.621486, -73.945646)	COLLEGE AVENUE	CAROLINE PLACE
7/3/12	21:50	MANHATTAN	10031	40.826446	-73.941542	(40.826446, -73.941542)	SAINT NICHOLAS PLACE	WEST 152 STREET
7/3/12	19:50	MANHATTAN	10029	40.794233	-73.946942	(40.794233, -73.946942)	EAST 108 STREET	PARK AVENUE
8/2/12	0:08	BROOKLYN	11211	40.714092	-73.942688	(40.714092, -73.942688)	METROPOLITAN AVENUE	HUMBOLDT STREET
8/3/12	16:00	STATEN ISLAND	10314	40.608832	-74.121169	(40.608832, -74.121169)	SCHUYLER LANE	SCHUYLER LANE
8/4/12	1:25	BROOKLYN	11220	40.647127	-74.015160	(40.647127, -74.015160)	SEAGATE BOULEVARD	VICTORY BOULEVARD
8/6/12	19:05	BROOKLYN	11204	40.618791	-73.983488	(40.618791, -73.983488)	51 STREET	3 AVENUE
8/8/12	2:09	BROOKLYN	11204	40.618791	-73.983488	(40.618791, -73.983488)	20 AVENUE	61 STREET
8/9/12	9:00	MANHATTAN	10022	40.761682	-73.974915	(40.761682, -73.974915)	WEST 55 STREET	5 AVENUE
8/9/12	22:30	MANHATTAN	10002	40.721438	-73.989136	(40.721438, -73.989136)	ALLEN STREET	STANTON STREET
8/10/12	22:25	BROOKLYN	11236	40.642041	-73.898829	(40.642041, -73.898829)	ROCKAWAY PARKWAY	AVENUE J
8/11/12	22:50	BROOKLYN	10314	40.791406	-73.935670	(40.791406, -73.935670)	ROCKAWAY PARKWAY	AVENUE J
8/11/12	12:18	BRONX	10474	40.812077	-73.886805	(40.812077, -73.886805)	RANDALL AVENUE	COSTER STREET
8/14/12	14:58	QUEENS	11377	40.762385	-73.903433	(40.762385, -73.903433)	GRANT HIGHWAY	WEST 170 STREET
8/20/12	8:31	QUEENS	10466	40.892477	-73.854492	(40.892477, -73.854492)	BOROUGH PLACE	BULOVA AVENUE
8/21/12	4:43	BRONX	10469	40.867550	-73.861503	(40.867550, -73.861503)	BARNES AVENUE	EAST 233 STREET
8/21/12	21:32	BRONX	11233	40.807687	-73.917023	(40.807687, -73.917023)	BRONXWOOD AVENUE	WILLIAMSBRIDGE ROAD
8/21/12	18:14	BROOKLYN	11233	40.807687	-73.917023	(40.807687, -73.917023)	EASTERN PARKWAY	SARATOGA AVENUE



Design Tip

How do you learn to make good visualizations?

...Make a lot of bad visualizations

HOW DOES YOUR OFFICE ANALYZE DATA?

Definition of Open Data

- Definition:
 - Open data is data that can be freely used, shared and built-on by anyone, anywhere, for any purpose
- Key Features
 - Availability and access
 - Reuse and redistribution
 - Universal participation

<http://blog.okfn.org/2013/10/03/defining-open-data/>

Open Data Benefits

- Transparency

IQuant NY

MAILING LIST RSS 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

MAY 11, 2016

The NYPD Was Systematically Ticketing Legally Parked Cars for Millions of Dollars a Year- Open Data Just Put an End to It

1705 Canton Avenue in Brooklyn, 273 Tickets, \$45,045 Legal.

270-05 76 Avenue in Queens, 256 Tickets (\$42,440) Legal.



<http://iquantny.tumblr.com/post/144197004989/the-nypd-was-systematically-ticketing-legally>

I Quant NY

[MAILING LIST](#) [RSS](#) [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](#)

MAY 11, 2016

The NYPD Was Systematically Ticketing Legally Parked Cars for Millions of Dollars a Year- Open Data Just Put an End to It

"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.

Yet, the majority of summonses written for this code violation were written by police officers. As a result, the department sent a training message to all officers clarifying the rule change and has communicated to commanders of precincts with the highest number of summonses, informing them of the issues within their command.

Thanks to this analysis and the availability of this open data, the department is also taking steps to digitally monitor these types of summonses to ensure that they are being issued correctly."

<http://iquantny.tumblr.com/post/144197004989/the-nypd-was-systematically-ticketing-legally>

Open Data Benefits

- Transparency
- Releasing social and commercial value

<https://healthyyout.com/>

<http://childcaredesk.com/>

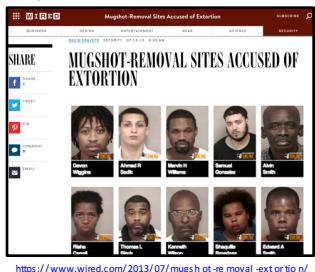
Open Data Benefits

- Transparency
- Releasing social and commercial value
- Participation and engagement



Open Data Concerns

- Privacy
 - Personally identifiable information (PII)



GAWKER

Public NYC Taxicab Database Lets You See How Celebrities Tip

J.K. Trotter
Filed to: DATA 10/23/14 1:00pm

138,411 ⚡ 18 ☆

BRADLEY COOPER

JULY 8, 2013 • 7:24 PM - 7:44 PM
376 GREENWICH ST. TO 13 BANK ST.
\$9.00 FARE • CASH; UNKNOWN TIP • ©SPLASH

<http://gawker.com/the-public-nyc-taxicab-database-that-accidentally-tracked-bradley-cooper-1646724546>

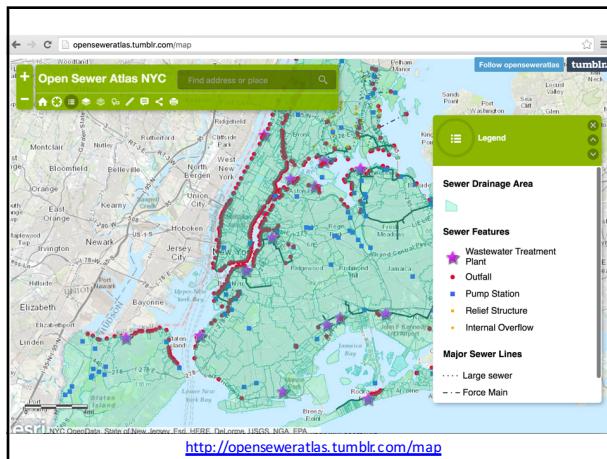
The Interdisciplinary Internet Institute

BLOG Using NYC Taxi Data To Identify Muslim Taxi Drivers

<http://www.theiii.org/index.php/997/using-nyc-taxi-data-to-identify-muslim-taxi-drivers/>

Open Data Concerns

- Privacy
 - Personally identifiable information (PII)
 - Mosaic Effect
- Confidentiality
- Security



LINKS TO OPEN DATA PORTALS

NYC Open Data Portal - <https://data.cityofnewyork.us/>

NYS Open Data Portal - <https://data.ny.gov>

US Federal Government Open Data Portal - <http://www.data.gov/>

Data for Exercise

Filter

Filter this dataset based on contents.

Created Date 01/01/2016 12:00:00 AM and 04/01/2016 12:00:00 AM

Complaint Type noise

Download

Download a copy of this dataset in a static format.

Download As
CSV
CSV for Excel
JSON
PDF
RDF
RSS
XLS
XLSX
XML

Exploratory Data Analysis

- Goal -> Discover patterns in the data
- Approach
 - Understand the context
 - Summarize fields
 - Use graphical representations of the data
 - Explore outliers

Tukey, J.W. (1977). Exploratory data analysis. Reading, MA: Addison-Wesley.

EXPLORING 311 NOISE COMPLAINTS

Question-Driven Analysis

- Goal -> Answer a specific problem or concern
- Approach
 - Have a question or problem in mind when analyzing data
 - “I need to know X”
 - Problem-focused discovery with the data

Question Driven Analysis

Vision Zero (dB)

Tasks:

- Given 311 noise complaint data, assist enforcement efforts by identifying community districts that have a high volume of noise complaints and the time frame enforcement resources should be deployed to combat the noise issue at its peak
- Identify the prevalent types of noise complaints in these areas to guide enforcement in each community district

NYC Community Districts

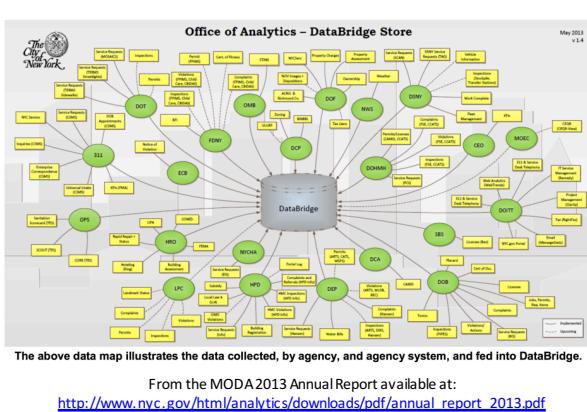


Analytical Resources

- Internal Agency Teams
 - Offices within your organization charged with performing analysis for internal or external stakeholders
- Mayor's Office Task Forces
 - Special inter-agency efforts around critical policy areas bringing together critical skills and experience in a subject area
- NYC Center for Innovation through Data Intelligence (CIDI)
 - Conducts inter-agency research to identify areas of service need in the City of New York
 - Collaborates with all Health and Human Service (HHS) agencies and other City partners to improve services
 - CIDI values the contextual interpretation of data and respects persons' confidentiality in its research activities
 - <http://www.nyc.gov/html/cidi/html/home/home.shtml>

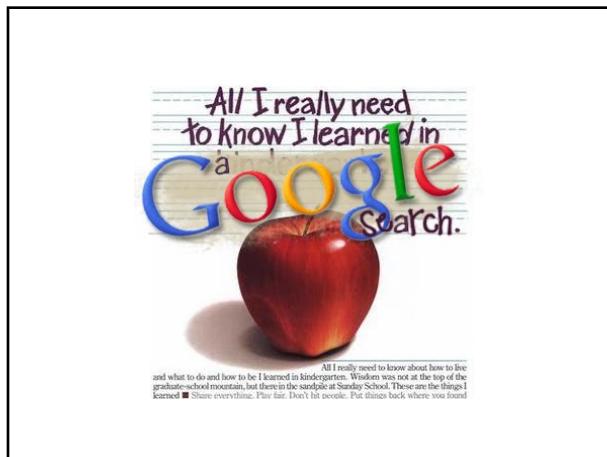
Analytical Resources

- Mayor's Office of Data Analytics (MODA)
 - New York City's civic intelligence center
 - Aggregating and analyzing data from across City agencies
 - More effectively address crime, public safety, and quality of life issues
 - Uses analytic tools to:
 - Prioritize risk more strategically
 - Deliver services more efficiently
 - Enforce laws more effectively
 - Increase transparency



What We've Covered

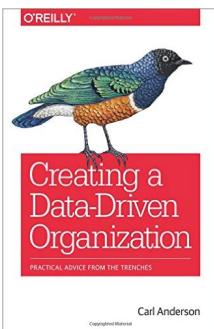
- 5 types of analysis
 - 4 concerns to be mindful of
 - Benefits of good analysis
 - 6 analytical steps
 - How to design charts and graphs
 - Definition of open data
 - Exploratory data analysis with 311 data



Technical Resources

- Stack Overflow
 - <http://stackoverflow.com/>
 - One of the best Q&A sites for technical questions of all kinds
 - Microsoft Office Support
 - <http://office.microsoft.com/en-us/support/>
 - Documentation on various MS Office products
 - Excel Tips
 - <http://excel.tips.net/>
 - Various tips and tricks for using Excel
 - Data Science Cheatsheet
 - <https://github.com/govex/DataScience>

Resources



Goals for the Course

- Discuss the data-driven decision making process
- Explore the role of managers and analysts in the decision making process
- Introduce useful terminology around data and the data analytics process
- Get some hands-on experience analyzing data

Key Takeaways for the Course

- Better understand the decision-making process with data
- Better understand the analytics process
- Better understand the value of data, particularly open data
- Better understand the role of analysts and managers in the decision-making process

Contact Information for MODA

Mayor's Office of Data Analytics

- Lindsay Mollineaux
 - Email: lmollineaux@cityhall.nyc.gov
- MODA website -
<http://www.nyc.gov/html/analytics/html/home/home.shtml>

Contact Information

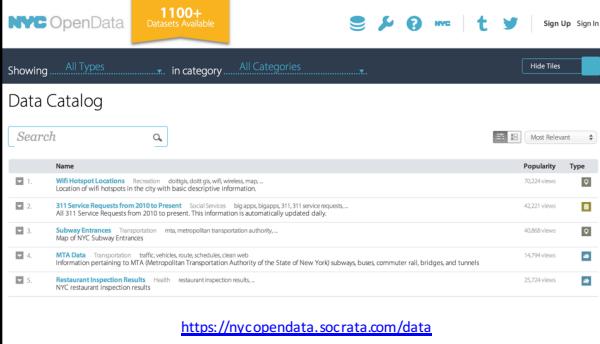
Instructor

- Name: Richard Dunks
- Email: richard@datapolitan.com
- Website: <http://www.datapolitan.com>
- Blog: <http://blog.datapolitan.com>
- Twitter: @rdunks1/@datapolitan

REFERENCE SLIDES

(Not presented in class)

NYC Open Data Portal



The screenshot shows the NYC Open Data Portal homepage. At the top, there's a banner with "NYC OpenData" and "1100+ Datasets Available". Below the banner, there are navigation links for "All Types", "In category", and "All Categories", along with "Hide Tiles" and search/filter options. The main area is titled "Data Catalog" and features a search bar. A list of datasets is displayed, each with a thumbnail, name, popularity (number of views), type, and a "View Details" button. The datasets listed are:

1. WiFi Hotspot Locations (Recreation) - 70,224 views
2. 311 Service Requests From 2010 to Present (Social Services) - 43,221 views
3. NYC Subway Map (Transportation) - 40,868 views
4. MTA Data (Transportation) - 14,794 views
5. Restaurant Inspection Results (Health) - 25,724 views

At the bottom of the catalog, there's a link to the URL: <https://nycopendata.socrata.com/data>.

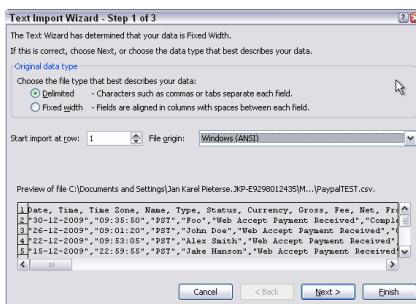
IMPORTING DATA

The following slides are for reference following the live demo in class

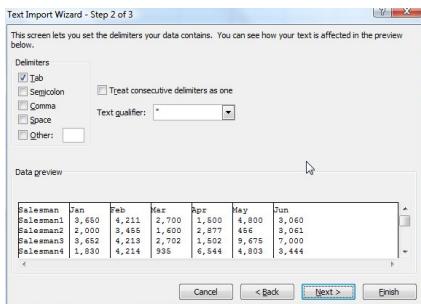
Importing Data



Importing Data



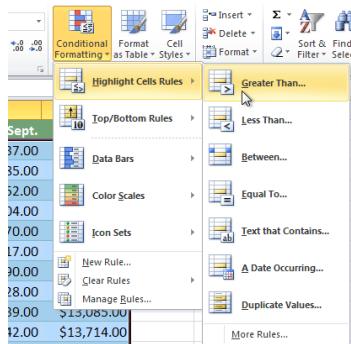
Importing Data



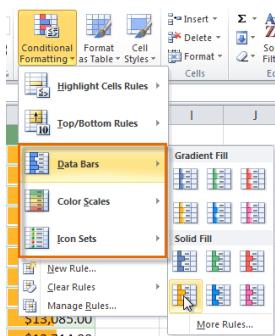
Conditional Formatting

- Format cells based on value or add content to cells that visually describe the content
- Great for quickly visualizing data
- Makes tables more “presentation-ready”

Conditional Formatting



Conditional Formatting

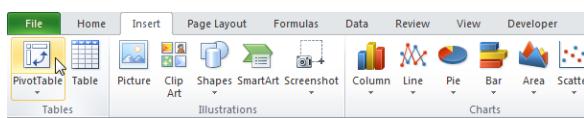


Conditional Formatting - Examples

\$3,863.00	\$1,117.00	\$8,237.00	\$8,690.00
\$9,355.00	\$1,100.00	\$10,185.00	\$18,749.00
\$6,702.00	\$2,116.00	\$13,452.00	\$8,046.00
\$4,415.00	\$1,089.00	\$4,404.00	\$20,114.00
\$3,863.00	\$1,117.00	\$8,237.00	\$8,690.00
\$9,355.00	\$1,100.00	\$10,185.00	\$18,749.00
\$6,702.00	\$2,116.00	\$13,452.00	\$8,046.00
\$4,415.00	\$1,089.00	\$4,404.00	\$20,114.00
⬇ \$3,863.00	⬇ \$1,117.00	⬇ \$8,237.00	⬇ \$8,690.00
⬇ \$9,355.00	⬇ \$1,100.00	⬇ \$10,185.00	⬆ \$18,749.00
⬇ \$6,702.00	⬇ \$2,116.00	↗ \$13,452.00	⬇ \$8,046.00
⬇ \$4,415.00	⬇ \$1,089.00	⬇ \$4,404.00	⬆ \$20,114.00

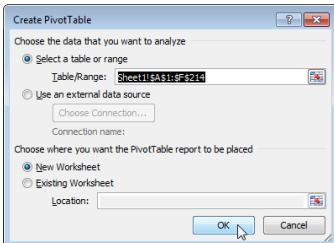
PivotTables

- What is a PivotTable?
 - A data summarization tool for quickly understanding and displaying the data you're analyzing
- How do I find it?



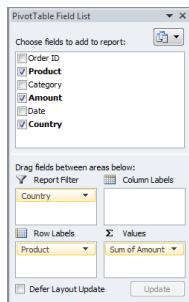
PivotTables

- Selecting range and destination



PivotTables

- Drag and drop fields to visualize
 - Row labels
 - Values
 - Filter
 - Column Labels



PivotTables

