



## Data Analytics for Managers

Instructors:

Richard Dunks and Julia Marden

Follow along at: <http://bit.ly/data-driven-gov>

## Datapolitan

Data Solutions for the Modern Metropolis

## Data Driven Culture



## Introductions

Turn to your neighbor and introduce yourself with the following:

- Who you are
- Where you work
- One data-related project you've worked on

You will introduce them to the class

## What to Expect for Today

- 9:15 – Welcome and Introductions
- 9:45 – Data Driven Government
- 10:30 – 15 min break
- 10:45 – Data Analytics and the Data Driven Culture
- 12:00 – Lunch
- 1:00 – Data Analytics Process and Open Data
- 1:55 – 5 min break
- 2:00 – Data Analytics Exercise
- 2:50 – 10 min break
- 3:00 – Data Analytics Exercise (continued)
- 4:30 – Dismissal

*All times approximate*

## Housekeeping

- We'll have one 15 minute break in the morning
  - We'll have an hour for lunch
  - We'll try for two breaks in the afternoon
  - Class will start promptly after breaks
  - Feel free to use the bathroom if you need during class
  - Please take any phone conversations into the hall to not disrupt the class

## 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.

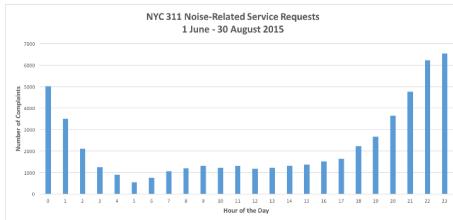


Retiring Intel CEO Paul Otellini

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

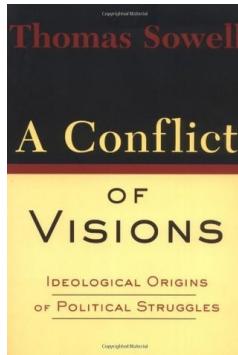
## The Value of Data

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



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*



Being data-driven doesn't mean



blindly following data.

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

### What is Analysis?

"Analysis is simply the pursuit of understanding, usually through detailed inspection or comparison"

- Carter Hewgley, Director of Analytics,  
Center for Government Excellence

---

---

---

---

---

---



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

---

---

---

---

---

---



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

---

---

---

---

---

---



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

---

---

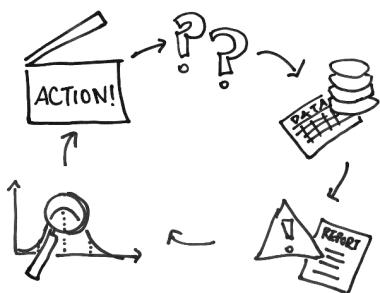
---

---

---

---

### It's Putting Them All Together



---

---

---

---

---

---

"If you do not know  
how to ask the  
right question,  
you discover  
nothing."  
W. Edward Deming



(c) iStockphoto / 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?

---



---



---



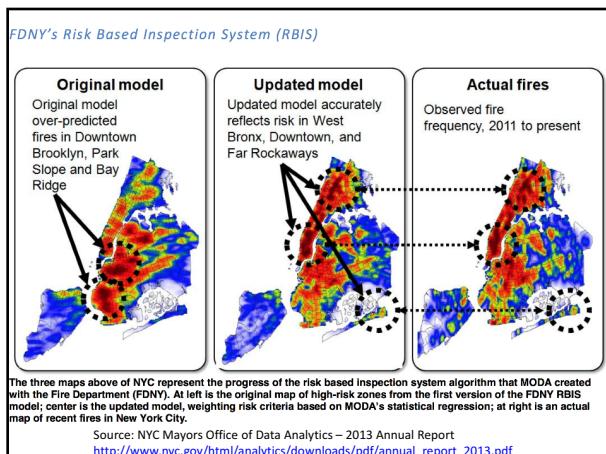
---



---



---




---



---



---



---



---



---

## 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?
  - “When a measure becomes a target, it ceases to be a good measure” - Charles Goodhart

---



---



---



---



---



---

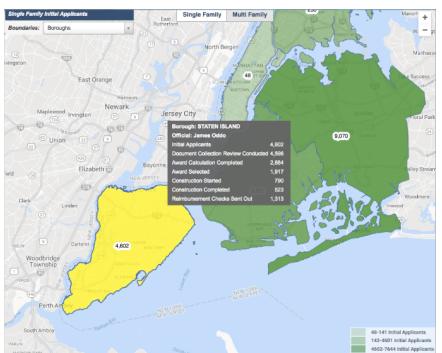


---



---

## Tracking Hurricane Sandy Relief Funds



Source: [http://www.slideshare.net/Leanalytics/startup-metrics-toronto-march-19](http://www.slideshare.net/Leananalytics/startup-metrics-toronto-march-19)

## Goodhart's Law

**"When a measure becomes a target, it ceases to be a good measure."**

- Charles Goodhart,  
Economist



<https://philebersole.wordpress.com/2013/01/10/goodharts-law-on-not-going-by-the-numbers/>

## 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

---



---



---



---



---



---

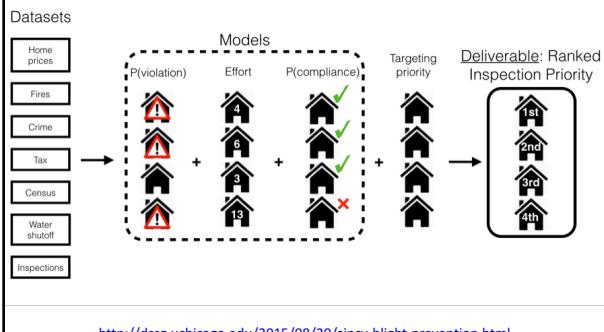


---



---

## Cincinnati Blight Prevention Model




---



---



---



---



---



---

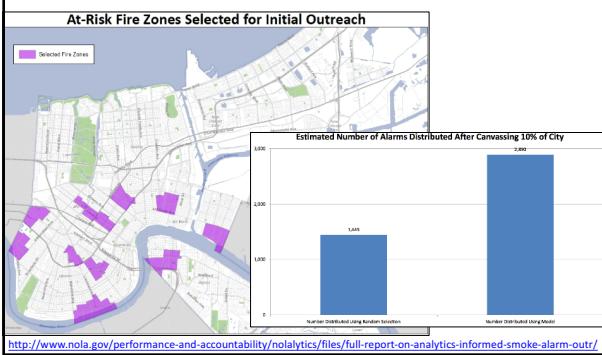


---



---

## New Orleans Smoke Alarm Targeted Outreach




---



---



---



---



---



---



---



---

## NOFD saves lives by targeted smoke alarm outreach

New Orleans Fire Department working hard to prevent fires: Chief Tim McConnell  
[www.nola.com](http://www.nola.com/politics/index.ssf?/base/politics&coll=1#4328111111111111)

Just two weeks ago, a family of 11 members, including a 2-month-old baby, escaped a disastrous fire in their Central City home. They were alerted by a smoke alarm at 3 a.m., fleeing before the smoke was bad enough to impede their escape. New Orleans firefighters had installed that smoke alarm in the home as part of this year's campaign. The truth is that I have never

<http://www.nola.com/politics/index.ssf?/base/politics&coll=1#4328111111111111>

**WHAT KIND OF ANALYSIS DOES YOUR OFFICE DO?**

**4 TYPES OF CONCERN 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

SUBSCRIBE LOG IN

528 COMMENTS

**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.

Email Share Tweet Save



[http://www.nytimes.com/2015/07/21/health/hipaa-s-use-as-code-of-silence-often-misinterprets-the-law.html?\\_r=0](http://www.nytimes.com/2015/07/21/health/hipaa-s-use-as-code-of-silence-often-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
- Public relations
- *Potential trap: not putting the necessary effort into something that will pay dividends to your agency and ultimately to the city as a whole*

---

---

---

---

---

---

### Benefits

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

---

---

---

---

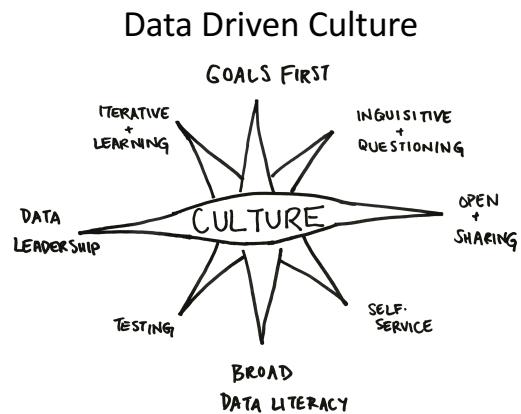
---

---

## 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

## BUILDING A DATA-DRIVEN 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

---

---

---

---

---

## 6 ANALYTICAL STEPS

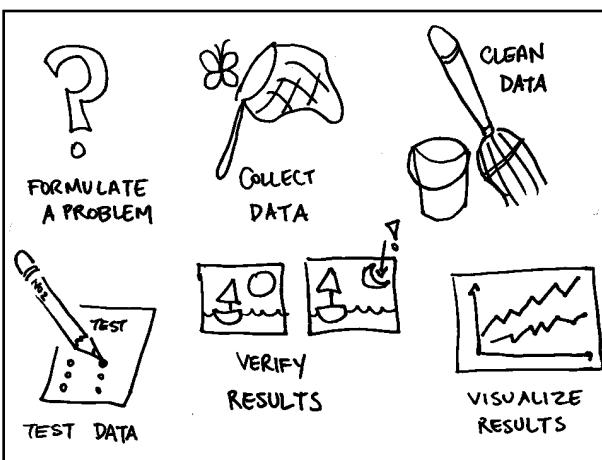
---

---

---

---

---



---

---

---

---

---

## 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

---

---

---

---

---

---

## 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

---



---



---



---



---



---

**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

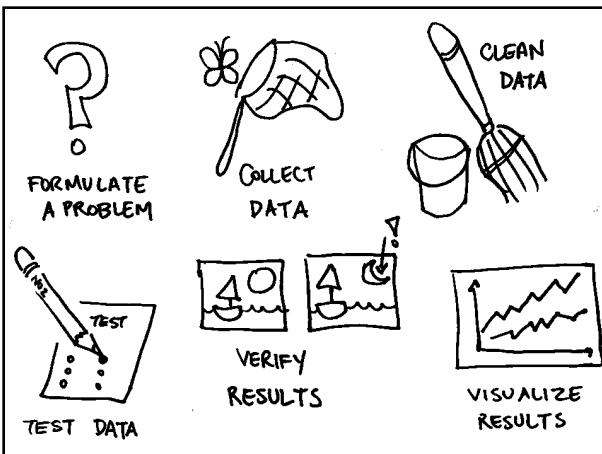
---

---

---

---

---



---

---

---

---

---

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

---

---

---

---

---

---

## 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

The screenshot shows a data portal interface with two main sections: 'Filter' and 'Download'.  
**Filter:** A dropdown menu titled 'Filter' with the sub-instruction 'Filter this dataset based on contents.' It contains two filter criteria:  

- Created Date** (dropdown) set to 'is between' with values '01/01/2016 12:00:00 AM' and '04/01/2016 12:00:00 AM'.
- Complaint Type** (dropdown) set to 'contains' with value 'noise'.

**Download:** A dropdown menu titled 'Download' with the sub-instruction 'Download a copy of this dataset in a static format.' It lists several options:  

- Download As:** CSV (selected), highlighted with a red box.
- JSON
- RDF
- RSS
- TSV for Excel
- XML

**Borough-Block-Lot:** 1001210001  
**Assessed Value:** \$70,771,500  
**Community District:** Manhattan 1  
**School District:** 2  
**Latitude:** 40.71306   **Longitude:** -74.0037

**Address:**  
**Census Tract:** 36061002900  
**Census Blocks:** 360610029004012  
**Coordinates:** X: 983210 Y: 199016

## 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.

---



---



---



---



---



---



---



---

## 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

---



---



---



---



---



---



---



---

## 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

- Elements of a data-driven culture
- Types of analysis
- The analytics process
- *What else?*

---



---



---



---



---



---



---

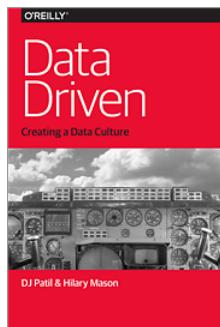
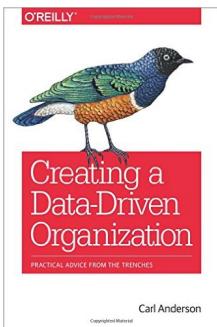


---

## 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/Data-Science>

## Resources



## Contact Information

### Instructors

- Richard Dunks
  - Email: [richard@datapoltan.com](mailto:richard@datapoltan.com)
  - Blog: <http://blog.datapolitan.com>
  - Twitter: @datapoltan
- Julia Marden
  - Email: [julia@tinypanther.com](mailto:julia@tinypanther.com)
  - Website: <http://tinypanther.com>
  - Twitter: @juliaem