



## Data Analytics for Managers

Instructor: Richard Dunks

25 May 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”

---



---



---



---



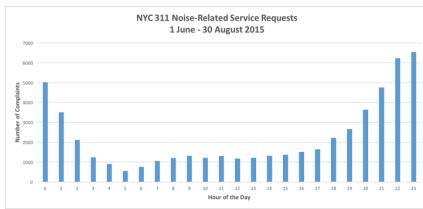
---



---

## 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 757

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=a2QIH2uz3p8>

---



---



---



---



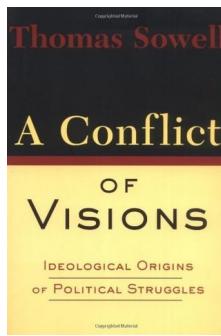
---



---

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*




---



---



---



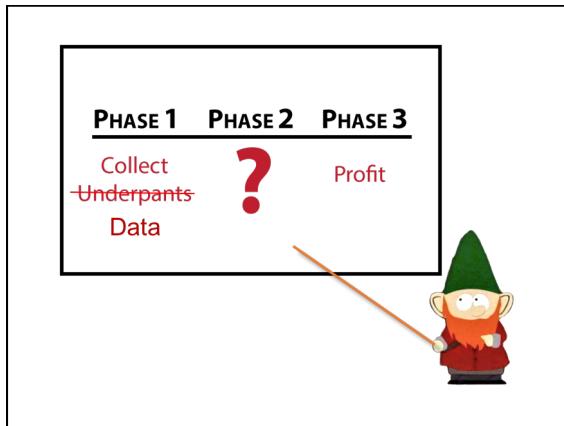
---



---



---



---

---

---

---

---



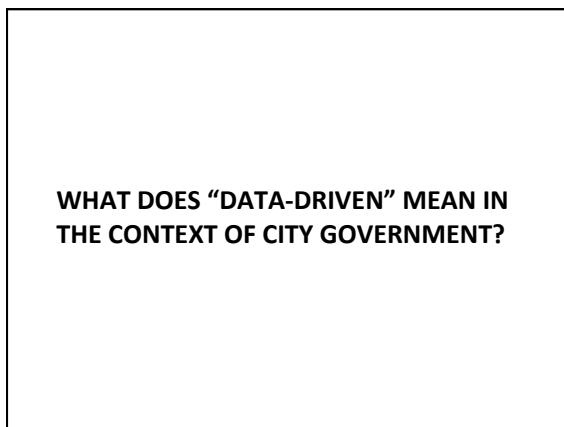
---

---

---

---

---



---

---

---

---

---

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.

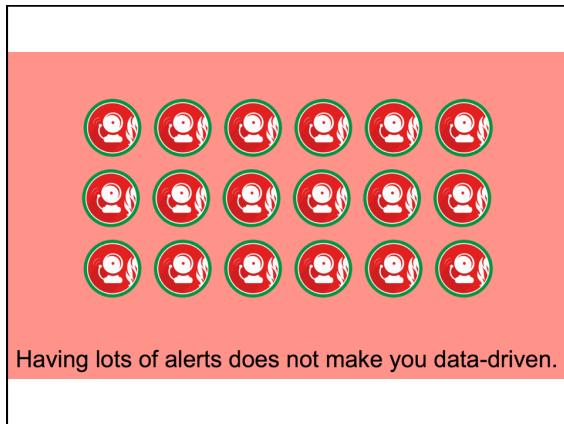
---

---

---

---

---



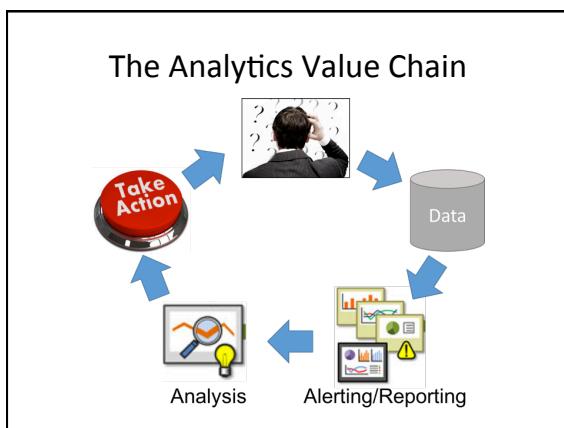
---

---

---

---

---



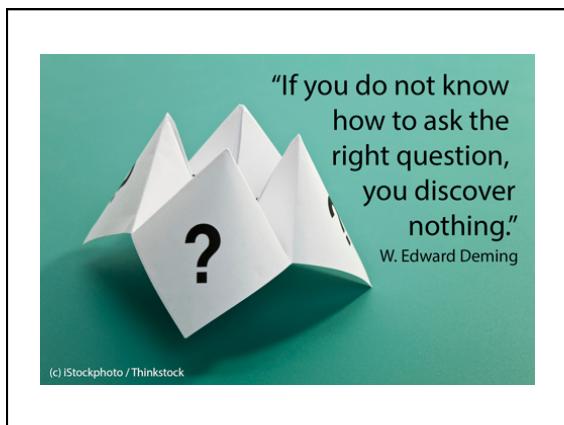
---

---

---

---

---



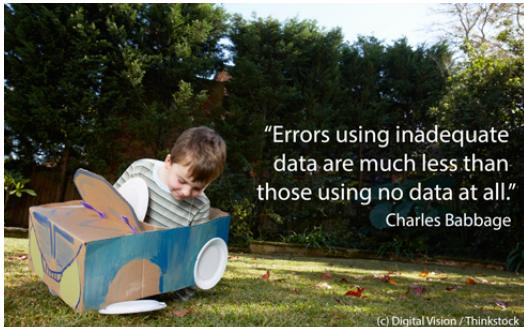
---

---

---

---

---



"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?

---

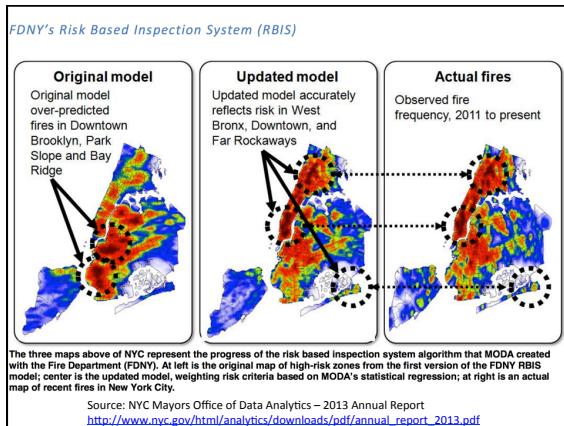
---

---

---

---

---



## 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](http://www1.nyc.gov/sandytracker)



Source: NYC Mayors Office of Data Analytics – 2013 Annual Report  
[http://www.nyc.gov/html/analytics/downloads/pdf/annual\\_report\\_2013.pdf](http://www.nyc.gov/html/analytics/downloads/pdf/annual_report_2013.pdf)

---



---



---



---



---



---



---



---



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

---



---



---



---

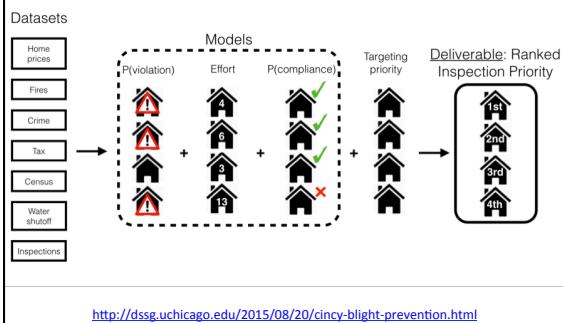


---



---

#### Cincinnati Blight Prevention Model




---



---



---



---

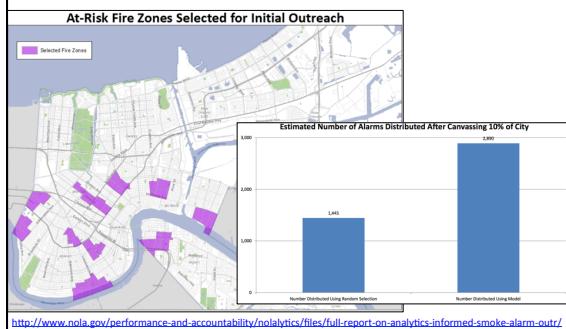


---



---

#### 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

---



---



---



---



---

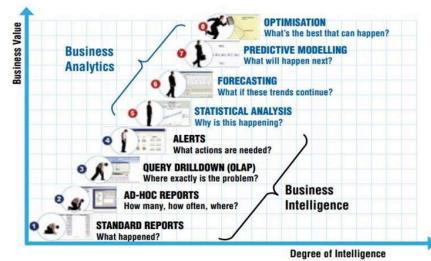


---



---

## Levels of Analytics



Source: [https://www.sas.com/news/sascom/analytics\\_levels.pdf](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/intel-chips-could-have-powered-first-iphone-ceo-otellini-says>

**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/Precendents
- *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/hipaas-use-as-code-of-silence-often-misinterprets-the-law.html?\\_r=0](http://www.nytimes.com/2015/07/21/health/hipaas-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
- *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

---



---



---



---



---



---

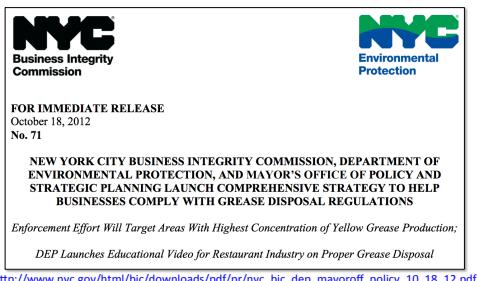
## Political Example – inBloom

- Public concern over the potential use of the data by 3<sup>rd</sup> 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/inbloom-shuts-down-amid-privacy-fears-over-student-data-tracking>

## Benefits

- Time, money, lives saved



## Benefits

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

**I Quant NY**

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.

MARCH 24, 2015

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

First, they have released a new fare card calculator which allows you to determine how to buy or refill a card to get even balance. 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.

\$1.00 Fee Applies

What amount do you want?

\$9.00 + \$0.00 BONUS  
\$19.00 + \$0.00 BONUS  
**\$27.25 + \$0.00 BONUS**  
\$39.00 + \$0.00 BONUS  
Other Amounts  
GO BACK CANCEL

<http://iquantny.tumblr.com/post/114470101209/i-quant-a-victory-mta-adds-new-button-for>

## Benefits

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

**I Quant NY**

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-saved-new>

**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

---



---



---



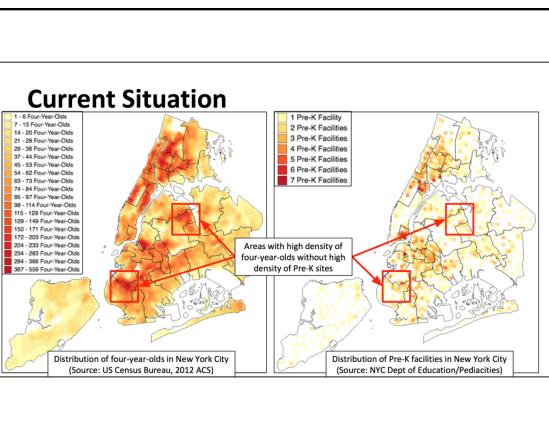
---



---



---




---



---



---



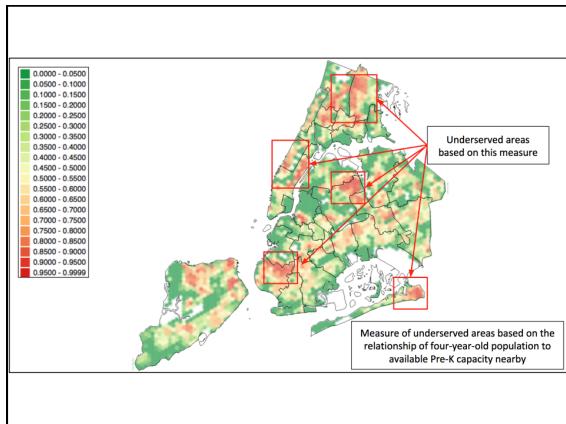
---



---



---




---



---



---



---



---



---




---



---



---



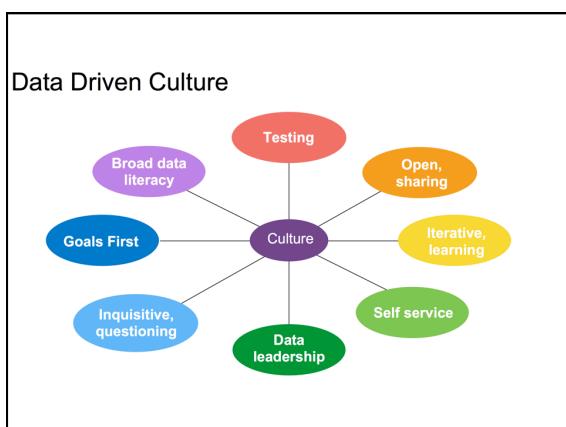
---



---



---




---



---



---



---



---



---



---



---

“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

---



---



---



---



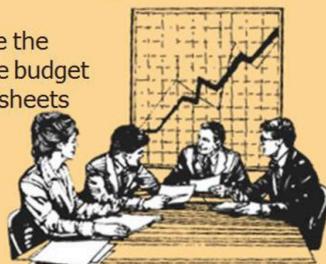
---



---

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...



a self service culture

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

---

---

---

---

---

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

---

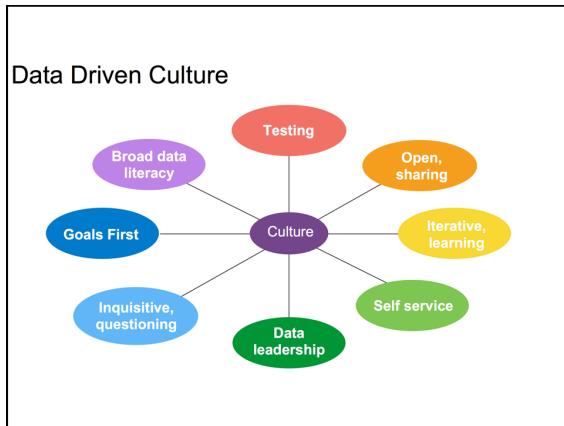
---

---

---

---

---




---



---



---



---



---



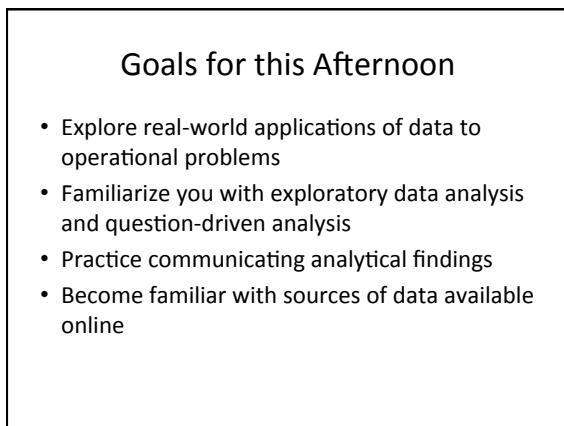
---



---



---




---



---



---



---



---



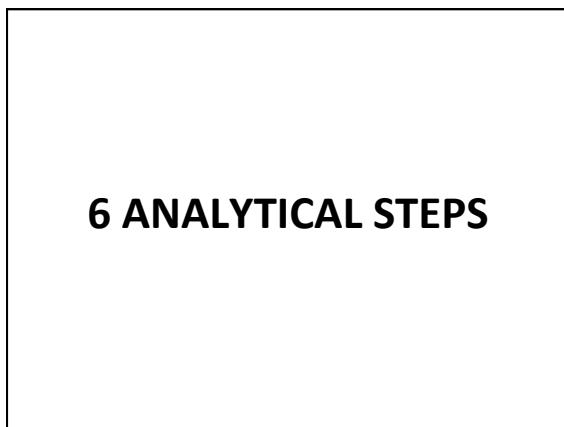
---



---



---




---



---



---



---



---



---



---



---

### 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. 17, 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



<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

---

---

---

---

---

---

---

---

## 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

---

---

---

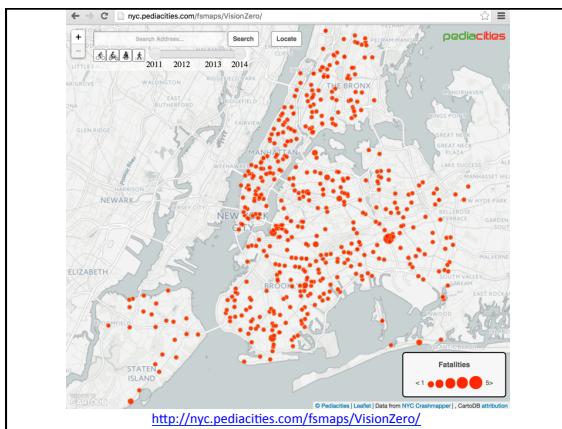
---

---

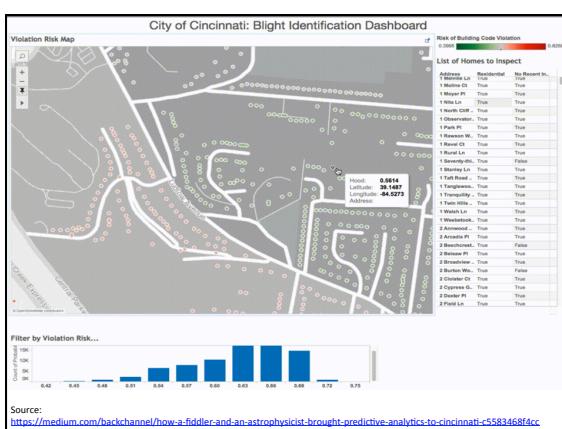
---

---

---



DATE	TIME	BOROUGH	ZIP CODE	LATITUDE	LONGITUDE	LOCATION	ON STREET NAME	CROSS STREET NAME
7/12/12	16:30	BROOKLYN	11216	40.685846	-73.947101	(40.685846, -73.947101)	MARYC AVENUE	MONROE STREET
7/12/12	16:30	BROOKLYN	11216	40.685846	-73.947101	(40.685846, -73.947101)	WILSON AVENUE	MONROE STREET
7/5/12	3:30	BROOKLYN	11235	40.6627363	-73.9571216	(40.6627363, -73.9571216)	BEDFORD AVENUE	STERLING STREET
7/8/12	21:54	BROOKLYN	11207	40.686646	-73.9026216	(40.686646, -73.9026216)	BUSHWICK AVENUE	CONWAY STREET
7/10/12	10:59	BROOKLYN	11204	40.725176	-73.8737187	(40.725176, -73.8737187)	OVERING STREET	EAST TREMONT AVENUE
7/13/12	2:18	BRONX	10467	40.8786453	-73.871598	(40.8786453, -73.871598)	WEBSTER AVENUE	EAST GUN HILL ROAD
7/14/12	21:20	MANHATTAN	10025	40.796567	-73.9703697	(40.796567, -73.9703697)	BROADWAY	WEST 99 STREET
7/14/12	1:21	BRONX	10456	40.887212	-73.937706	(40.887212, -73.937706)	GRAND CONCOURSE	EAST 137 STREET
7/14/12	2:20	BRONX	10456	40.887212	-73.937706	(40.887212, -73.937706)	GRAND CONCOURSE	EAST 137 STREET
7/16/12	10:55	BROOKLYN	11230	40.69674	-73.9552332	(40.69674, -73.9552332)	EAST 8 STREET	AVENUE P
7/18/12	22:51	QUEENS	11104	40.7390635	-73.9265179	(40.7390635, -73.9265179)	GREENPOINT AVENUE	39 PLACE
7/21/12	2:08	QUEENS	11148	40.6857978	-73.723999	(40.6857978, -73.723999)	MURK BOULEVARD	VICTORIA ROAD
7/22/12	20:00	BROOKLYN	11203	40.6857978	-73.723999	(40.6857978, -73.723999)	MURK BOULEVARD	DUMBO
7/23/12	2:08	QUEENS	11145	40.6972049	-73.833045	(40.6972049, -73.833045)	AVENUE 1	VAN WYCK EXPRESSWAY
7/28/12	0:45	STATEN ISLAND	10318	40.621429	-74.1264267	(40.621429, -74.1264267)	COLLEGE AVENUE	CAROLINA PLACE
7/30/12	21:52	MANHATTAN	10003	40.828446	-73.9451422	(40.828446, -73.9451422)	SAIN NICOLAS PLACE	WEST 152 STREET
7/30/12	1:00	MANHATTAN	10003	40.828446	-73.9451422	(40.828446, -73.9451422)	EAST 108 STREET	PARK AVENUE
8/1/12	10:08	BROOKLYN	11231	40.6742333	-73.9452462	(40.6742333, -73.9452462)	WILSON AVENUE	HUMBOLDT STREET
8/3/12	16:00	STATEN ISLAND	10318	40.6088332	-74.1211698	(40.6088332, -74.1211698)	MANDY ROAD	SCHMIDTS LANE
8/4/12	12:28						BAY STREET	VICTORY BOULEVARD
8/6/12	1:20	BROOKLYN	11204	40.6471257	-74.031300	(40.6471257, -74.031300)	51 STREET	3 AVENUE
8/8/12	2:09	BROOKLYN	11204	40.6471257	-74.031300	(40.6471257, -74.031300)	51 STREET	41 STREET
8/9/12	9:00	MANHATTAN	10202	40.7616692	-73.9749156	(40.7616692, -73.9749156)	WEST 55 STREET	5 AVENUE
8/9/12	22:30	MANHATTAN	10002	40.7216482	-73.9891302	(40.7216482, -73.9891302)	ALLEN STREET	STANTON STREET
8/10/12	20:25	BROOKLYN	11238	40.6742331	-73.9888296	(40.6742331, -73.9888296)	ROCKAWAY PARKWAY	AVENUE J
8/11/12	1:20	BRONX	10465	40.8714008	-73.9452069	(40.8714008, -73.9452069)	GRANT AVENUE	AVENUE J
8/11/12	1:20	BRONX	10474	40.8120773	-73.8868051	(40.8120773, -73.8868051)	RANDALL AVENUE	COSTER STREET
8/14/12	14:58						GRANT AVENUE	WEST 170 STREET
8/20/12	1:20	QUEENS	11177	40.6283853	-73.9043034	(40.6283853, -73.9043034)	BROOKLYN PARK	BROOKLYN PARK
8/21/12	2:12	BRONX	10465	40.8755074	-73.8545023	(40.8755074, -73.8545023)	BRUNNIE AVENUE	EAST 123 STREET
8/21/12	2:12	BRONX	10469	40.8755031	-73.8615031	(40.8755031, -73.8615031)	BRONWOOD AVENUE	WILLIAMSBRIDGE ROAD
8/22/12	18:14	BROOKLYN	11233	40.6706879	-73.9170232	(40.6706879, -73.9170232)	STEINERN 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
- Releasing social and commercial value

---



---



---



---



---



---

*I Quant NY*

MAILING LIST RSS ARCHIVE

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-saved-new>

---



---



---



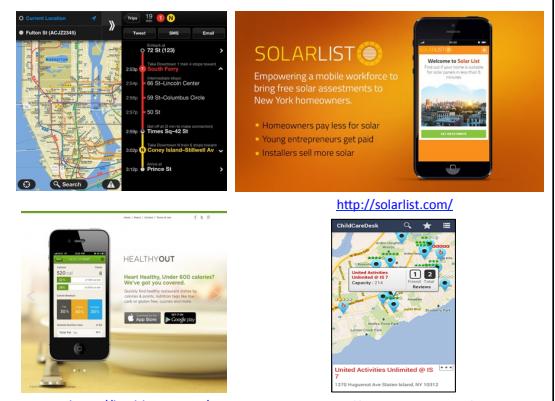
---



---



---



<http://solarlist.com/>

<https://healthyout.com/>

<http://childcaredesk.com/>

---



---



---



---



---



---

## Open Data Benefits

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



<https://www.votinginfoproject.org/>

---



---



---



---



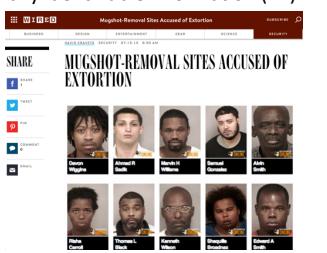
---



---

## Open Data Concerns

- Privacy
  - Personally identifiable information (PII)




---



---



---



---



---



---

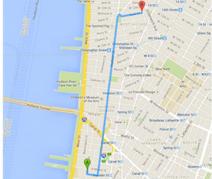


**GAWKER**

## Public NYC Taxicab Database Lets You See How Celebrities Tip

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

138,411 18 5\*

BRADLEY COOPER

JULY 8, 2013 - 7:34 PM - 7:42 PM  
374 GREENWICH ST TO 12 BANK ST.  
\$9.0 FARE • CASH: UNKNOWN TIP • SPLASH

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

The Interdisciplinary Internet Institute

About Contributors Join Us Contact

BLOG

### Using NYC Taxi Data To Identify Muslim Taxi Drivers

© 21 JAN, 2015 | BY ANNA BERLE

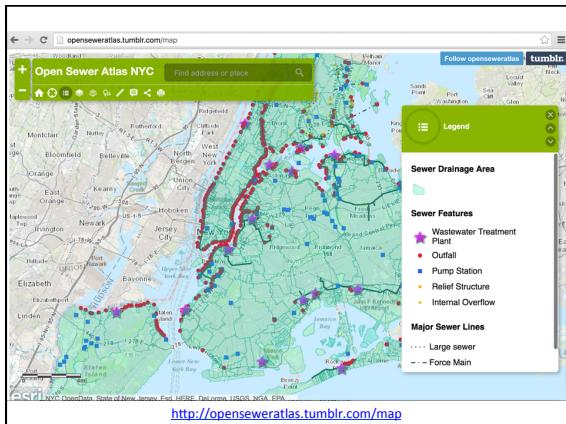
Remember that NYC Taxi data set that allowed you to see who visited a gentlemen's club and which celebrity took a taxi where? Reddit user ulman now seems to have found a way to distinguish Muslim taxi drivers from the rest. He explains how:

Since Islam instructs followers to pray 5x daily at specific times, I wondered if one could identify devout Muslims solely from their trip data. For drivers that do pray regularly, there are surely different times a place to park, wait, and pray at the exact time, but in many cases based near prayer times. In detail, I plotted the fare data for each driver. Each image shows fares for one cab in 2013. Yellow-active fare (carrying passengers). A minute is 1 pixel wide; a day is 2 pixels tall. Blue stripes indicate the 5 daily prayer start times which vary with the sun's position throughout the year.

<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



<http://openseweratlas.tumblr.com/map>

## 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 -> is between  
 01/01/2016 12:00:00 AM and  
 04/01/2016 12:00:00 AM  
 and

Complaint Type -> contains  
 noise  
 and

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

---



---



---



---



---



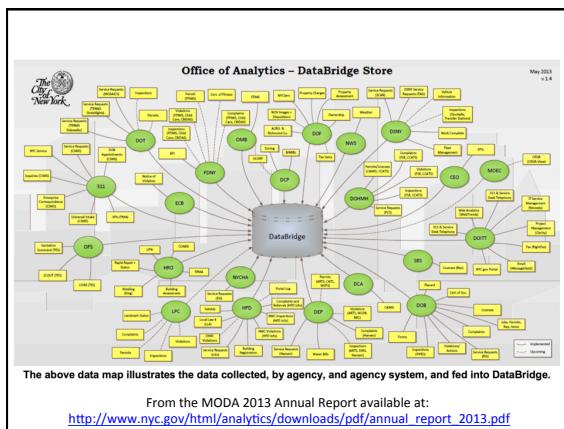
---



---



---




---



---



---



---



---



---



---



---

## What We've Covered

- 5 types of analysis
- 4 concerns to be mindful of
- Benefits of good analysis
- 6 analytical steps
- 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

---



---



---



---



---

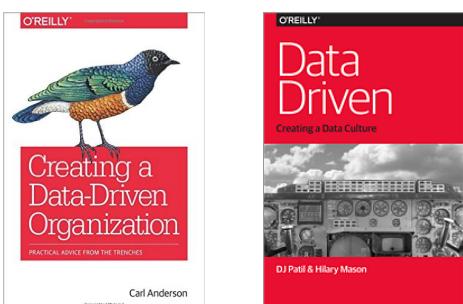


---



---

## 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](mailto: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](mailto:richard@datapolitan.com)
- Website: <http://www.datapolitan.com>
- Blog: <http://blog.datapolitan.com>
- Twitter: @rdunks1/@datapolitan

---

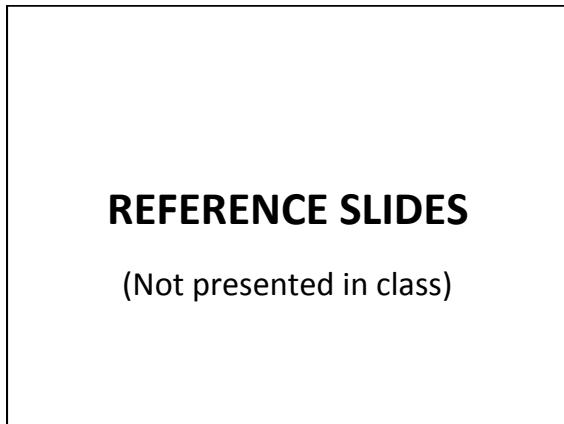
---

---

---

---

---




---



---



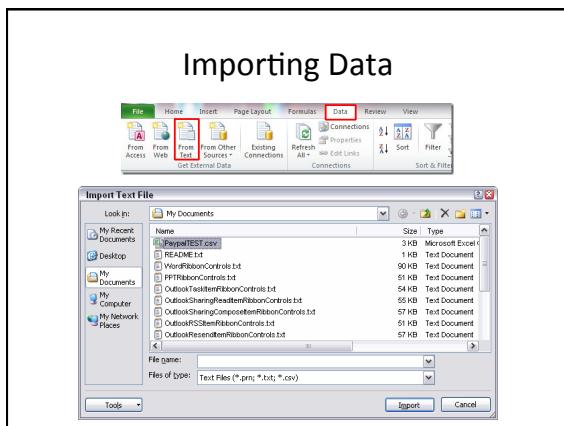
---



---



---




---



---



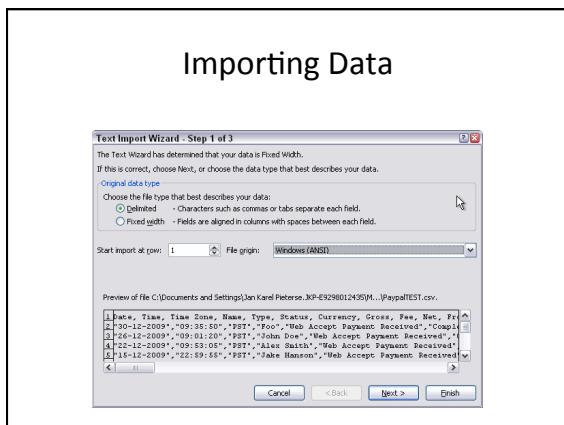
---



---



---




---



---



---

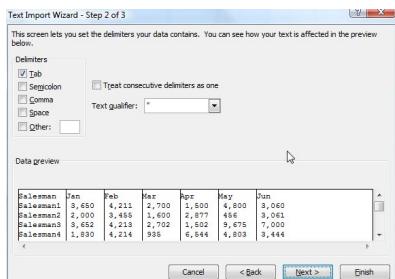


---



---

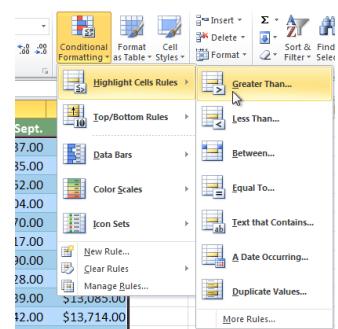
## Importing Data

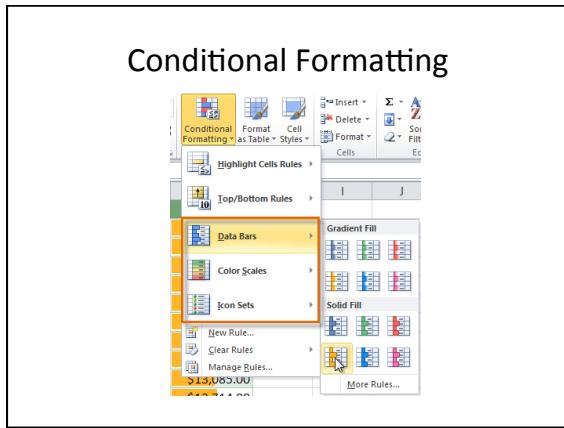


## Conditional Formatting

- 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 - 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?

The screenshot shows the Microsoft Excel ribbon with the 'Insert' tab selected. In the 'Tables' group, the 'PivotTable' icon is highlighted with a yellow box.

---

---

---

---

---

## PivotTables

- Selecting range and destination

The dialog box is titled 'Create PivotTable'. It has two main sections: 'Choose the data that you want to analyze' and 'Choose where you want the PivotTable report to be placed'. In the first section, 'Select a table or range' is selected, and the 'TableRange' dropdown shows 'Sheet1!\$A\$1:\$F\$21'. Below it, there's a 'Choose Connection...' button and a 'Connection name:' dropdown. In the second section, 'New Worksheet' is selected, and the 'Location:' dropdown is empty. At the bottom are 'OK' and 'Cancel' buttons.

## PivotTables

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

The window is titled 'PivotTable Field List'. It has two main sections: 'Choose fields to add to report:' and 'Drag fields between areas below:'. In the first section, several fields are listed with checkboxes: Order ID (unchecked), Product (checked), Category (unchecked), Amount (unchecked), Date (unchecked), and Country (checked). In the second section, 'Report Filter' is set to 'Country', 'Column Labels' is empty, 'Row Labels' is set to 'Product', and 'Values' is set to 'Sum of Amount'. At the bottom are 'Update' and 'Defer Layout Update' buttons.

## PivotTables

The context menu on the left includes options like Copy, Format Cells..., Refresh, Sort, Remove 'Sum of Sales', Summarize Values By, Show Values As, Show Details, Value Field Settings..., PivotTable Options..., and Hide Field List. The 'Value Field Settings...' option is highlighted. To its right, a 'Value Field Settings' dialog box is open. It shows 'Source Name: Sales' and 'Custom Name: Count of Sales'. Under 'Summarize Values By', it says 'Show Values As' and 'Summarize value field by'. A dropdown menu lists 'Sum', 'Average', 'Max', 'Min', and 'Product', with 'Sum' currently selected. At the bottom are 'Number Format', 'OK', and 'Cancel' buttons.