

Data**politan**

Data Solutions for the Modern Metropolis

Data Analytics for Managers

Instructor: Richard Dunks

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

See the code at: <http://bit.ly/data-driven-gov-code>

Welcome

A Few Ground Rules

- Step up, step back
- Be curious and ask questions!
- Assume noble regard and positive intent
- Respect multiple perspectives
- Listen deeply
- Be present (phone, email, social media, etc.)

Introduce Yourself to Your Neighbor

- Who are you?
- Where do you work?
- What has been the proudest moment in your job?

What to Expect Today

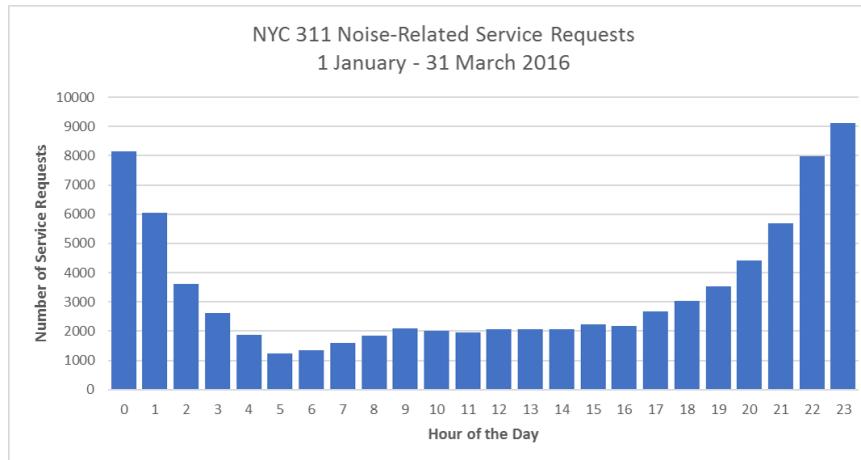
- 9:40 – Data Analytics 101
- 10:00 - Introduction to Problem Ideation
- 10:30 – 15 min break
- 10:45 – Process Mapping and the Types of Analysis
- 12:00 – Lunch
- 1:00 – Overview of Open Data
- 2:00 – Data Analytics Exercise
- 2:30 – 15 min break
- 2:45 – Data Analytics Exercise (continued)
- 4:30 – Dismissal

Housekeeping

- We'll have one 15 minute break in the morning
- We'll have an hour for lunch
- We'll have a 15 minute break 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

The Value of Data

- Data tells a story about something that's happened
- Can describe what happened directly or indirectly



Are All Data Points Created Equal?

“ 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*

Data Driven Decisions Require Humans



Image Credit: 100 lion, CC BY-SA 4.0, via Wikimedia Commons

How Intel lost out on the contract of a lifetime

‘ Data is only as valuable as the decisions it enables.

-Ion Stoica

‘ If data is enabling important decisions, then the data is important too

What is Analysis?

‘ “Analysis is simply the pursuit of understanding, usually through detailed inspection or comparison”

- Carter Hewgley, (Former) Director of Analytics, Center for Government Excellence



What Analysis Isn't



Writing a report

Creating a dashboard

<http://localhost:8895/#1>

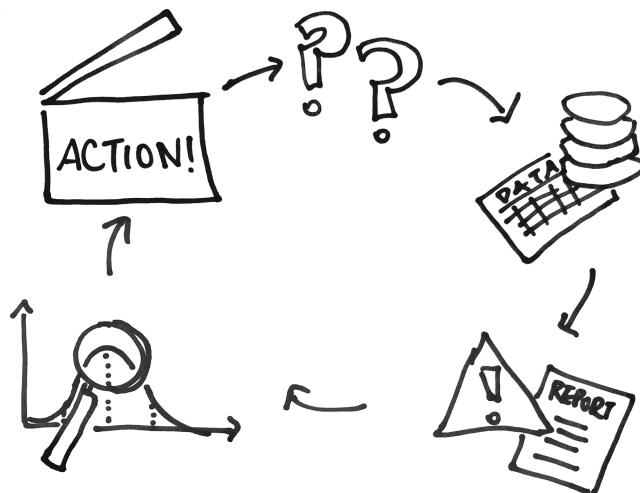
Page 15 of 122

Generating an alert



Image Credit: Tokyoship, CC BY-SA 3.0, via Wikimedia Commons

It's Putting Them All Together

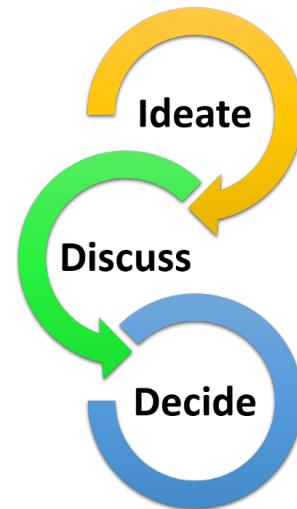


“If you do not know how to ask the right question, you discover nothing.”

- W. Edward Deming

Our Method For Generating Ideas

- **Ideate** - On your own, generate at least 3 ideas (ideally more), each on their own Post-It Note
- **Discuss** - Review the ideas generated
- **Decide** - Come to a consensus as a group



Exercise - Reducing Noise Complaints in NYC

- Between 1 Jan and 30 Sept 2017, there were an average of 1,271 noise-related 311 service requests a day
- The same period in 2016 had an average of 1,180 noise-related 311 service requests a day
- You've been tasked with decreasing noise complaints in the city
- **What questions would you ask to kick off a data analysis?**

15 MIN BREAK

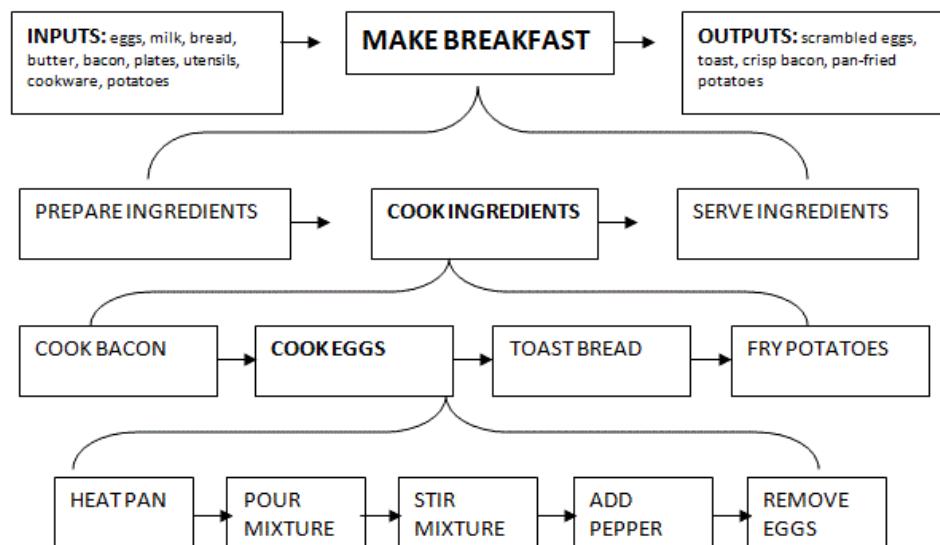


Process Mapping (Our method)

Process Mapping

- Allows you to identify and strategize for key steps in your analysis
- Helps sequence tasks and identify gaps in understanding
- Provides a basis for documenting work

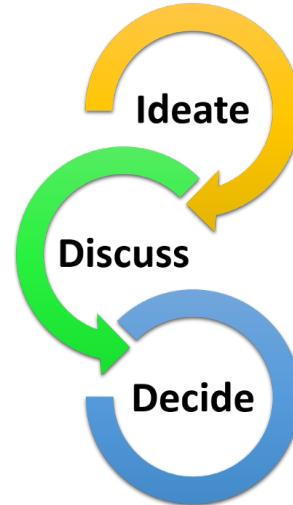
Process Mapping



How to Create a Process Map

- Identify the key challenge
- Identify the outcome
- Identify key way to validate the outcome (outputs)

How do we know we've got it right?

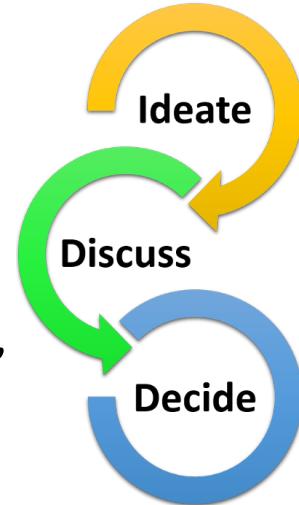


Outcomes vs Outputs

- Outcomes are the larger benefits and/or achievements you're trying to realize (happiness, health, well-being, etc.)
- Outputs are the tangible parts of your outcome (survey responses, measured results, etc.)
- Outputs enable us to find outcomes
- Without outcomes, there is no need for outputs

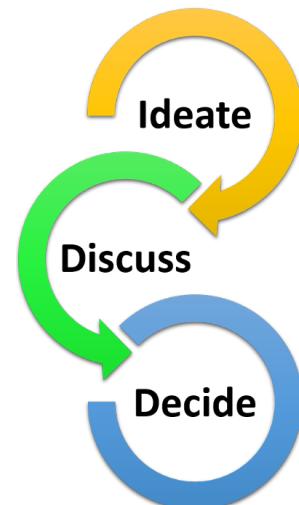
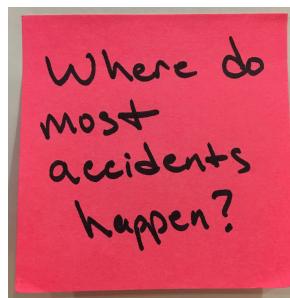
How to Create a Process Map

- Identify the key challenge
- Identify the outcome
- Identify key way to validate the outcome (outputs)
How do we know we've got it right?
- Identify the key inputs (data, partners, etc.)
- Sequence the key questions to turn inputs into outputs



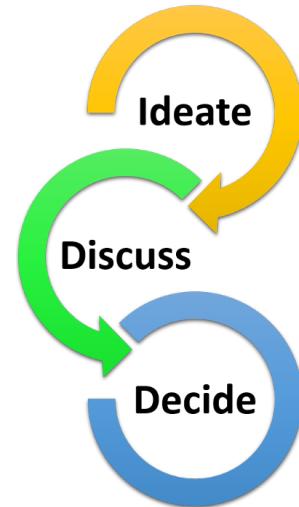
Key Tips

- Place each step on a Post-It Note



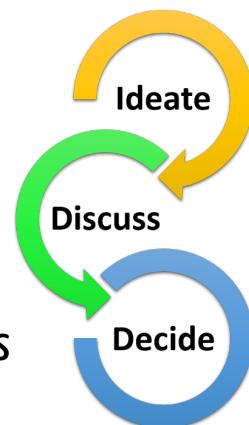
Key Tips

- Place each step on a Post-It Note
- Order and reorder as necessary
- Some steps will need to be broken down



Reducing Noise Complaints

- Identify the key question/challenge
- Identify the key outcome
- Identify the outcome measures
- Identify the key inputs (data, partners, etc.)
- Sequence the key questions to turn inputs into outputs

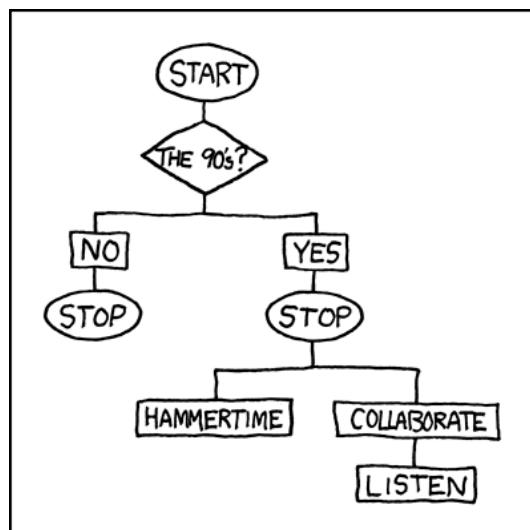


WRAP-UP

<http://localhost:8895/#1>

Page 31 of 122

LUNCH

[Source](#)<http://localhost:8895/#1>

Page 32 of 122



WELCOME BACK!



Link to NYC Open Data Portal for Exercise

[NYC Open Data Portal](#)

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
options

Complaint Type contains
noise
and
options

Download

Download a copy of this dataset in a static format

Download As

CSV
CSV for Excel
JSON
RDF
RSS
TSV for Excel
XML

[Click to download if you have problems](#)

5 Data Analytics Tasks

1. Sorting
2. Filtering
3. Aggregating (PivotTable)
4. Manipulating
5. Visualizing

1. Sorting

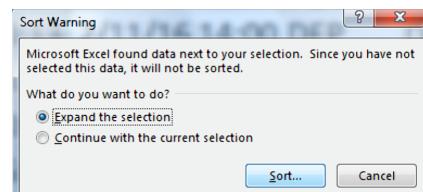
- Reorganize rows in a dataset based on the values in a column
- Can sort on multiple columns

Sorting by Date

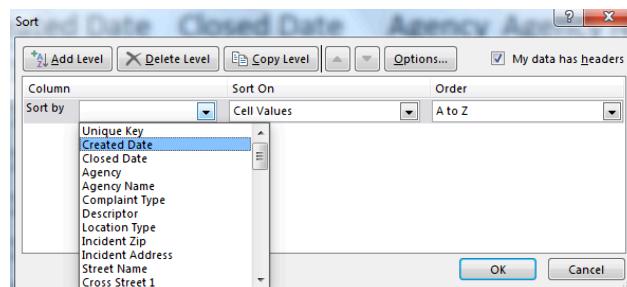
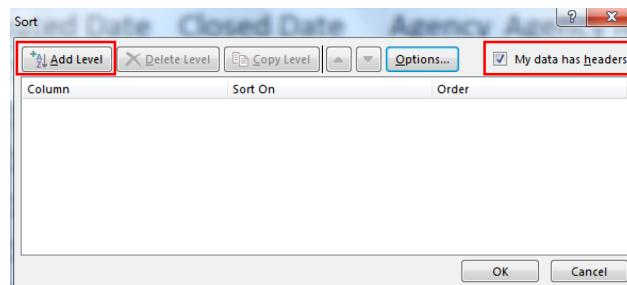
	A	B	C
1	Unique Key	Created Date	Closed Date
2	32576232	2/1/16 10:13	2/11/16 14:00
3	32884992	3/11/16 13:48	3/12/16 8:00
4	32445685	1/18/16 10:56	1/20/16 8:00
5	32816604	3/2/16 15:17	3/3/16 8:00
6	32480740	1/22/16 10:19	1/24/16 20:00
7	32518960	1/25/16 7:35	1/28/16 15:45



	A	B	C
1	Unique Key	Created Date	Closed Date
2	32305299	1/1/16 0:00	1/1/16 1:57
3	32310343	1/1/16 0:00	1/1/16 3:12
4	32308578	1/1/16 0:02	1/1/16 23:35
5	32305983	1/1/16 0:03	1/1/16 3:24
6	32305208	1/1/16 0:03	1/1/16 2:43
7	32309484	1/1/16 0:04	1/1/16 0:28

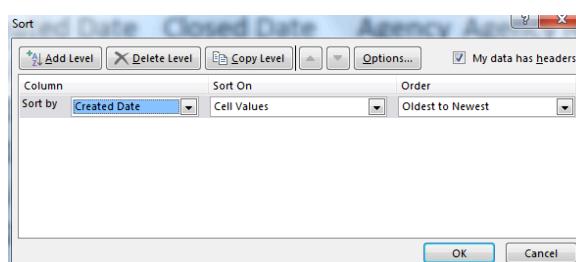


Sorting by Date


<http://localhost:8895/#1>

Page 39 of 122

Sorting by Date



	A	B	C
1	Unique Key	Created Date	Closed Date
2	32305299	1/16 0:00	1/16 1:57
3	32310343	1/16 0:00	1/16 3:12
4	32308578	1/16 0:02	1/16 23:35
5	32305983	1/16 0:03	1/16 3:24
6	32305208	1/16 0:03	1/16 2:43
7	32309484	1/16 0:04	1/16 0:28

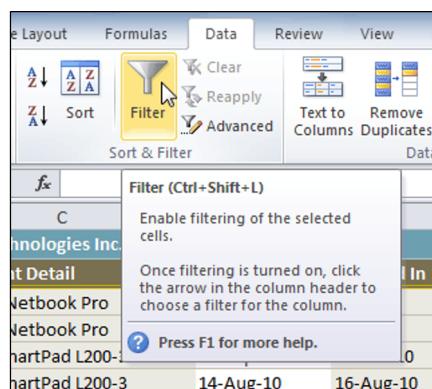
<http://localhost:8895/#1>

Page 40 of 122

2. Filtering

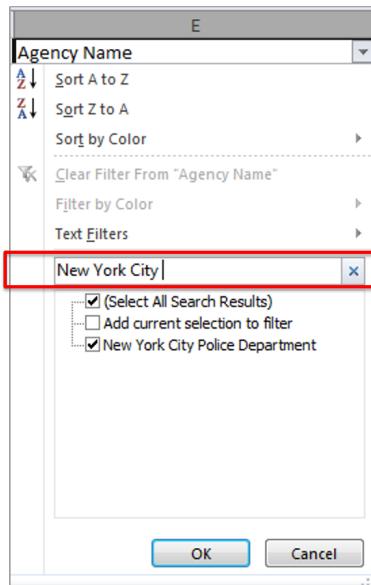
- Only show rows that contain some value
- Can filter by multiple values
- Can filter by values in multiple columns

Filtering by Agency Name



	A	B	C	D	E	F
1	Unique Key	Created Date	Closed Date	Agency Name	Complaint Type	

Filtering by Agency Name

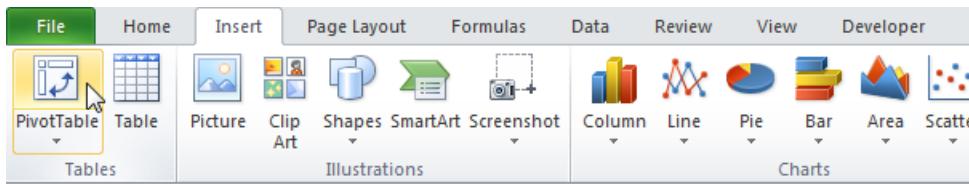


3. Aggregating Data

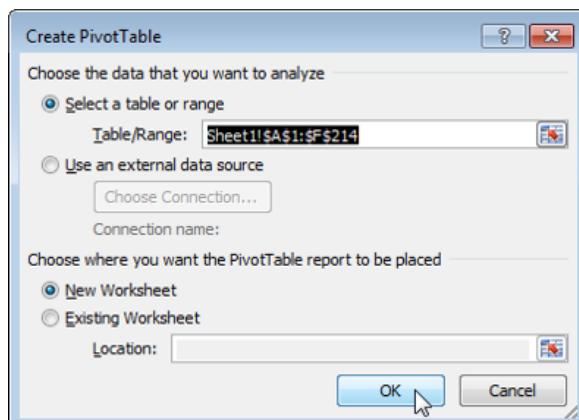
- Trends only become clear in aggregate
- Often where you discover the "so what"
- Aggregating data meaningfully can be tricky

PivotTables

- A data summarization tool
- Useful to quickly understand data
- Can use to graph data totals



Creating a PivotTable

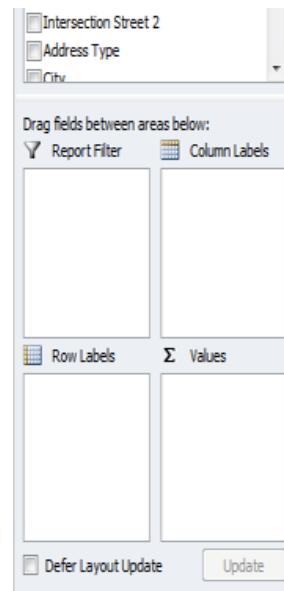


- Should default to all your data unless you have any cells selected

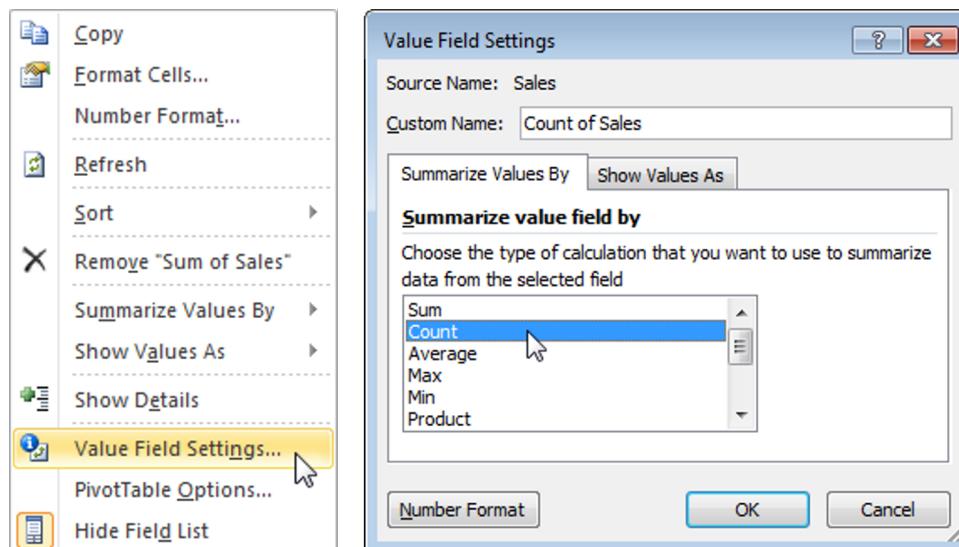
Creating a PivotTable

Drag and drop fields to visualize

- Row labels
- Values
- Filter
- Column Labels



Creating a PivotTable



4. Manipulating Data (In a good way...)

- Sometimes available categories don't make sense
- Values may not be in the format you need (or have mistakes)
- You always want to have a clean copy of the data to go back to
- Best to keep track of what you've done

Extracting Hour From Timestamp

A	B	C
Unique Key	Created Date	hour
30195273	3/18/15 2:12	=HOUR(B2)
30203057	3/18/15 2:00	
30197320	3/18/15 1:58	
30194112	3/18/15 1:37	
30202379	3/18/15 1:36	
30199506	3/18/15 1:28	

HOUR function

This article describes the formula syntax and usage of the **HOUR**function in Microsoft Excel.

Description

Returns the hour of a time value. The hour is given as an integer, ranging from 0 (12:00 A.M.) to 23 (11:00 P.M.).

Syntax

`HOUR(serial_number)`

The HOUR function syntax has the following arguments:

- **Serial_number** Required. The time that contains the hour you want to find. Times may be entered as text strings within quotation marks (for example, "6:45 PM"), as decimal numbers (for example, 0.78125, which represents 6:45 PM), or as results of other formulas or functions (for example, `TIMEVALUE("6:45 PM")`).

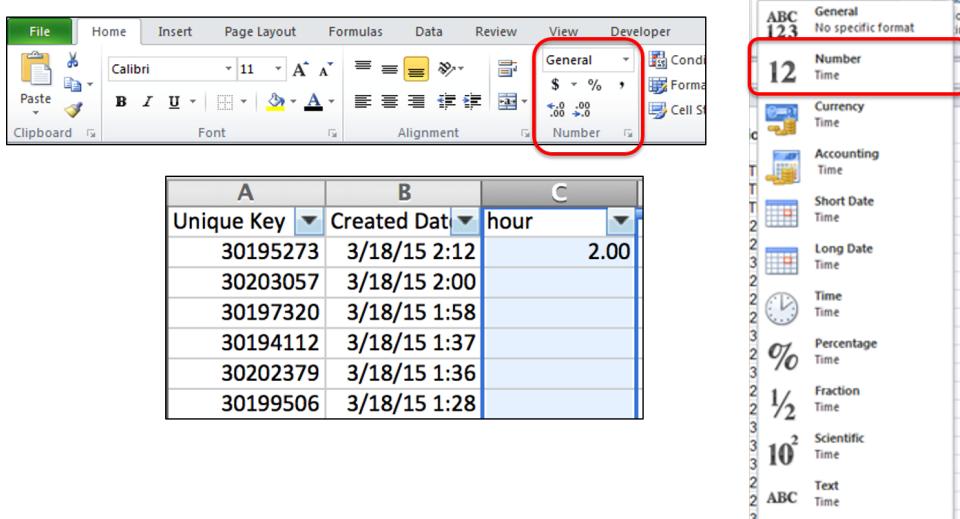
Extracting Hour From Timestamp

Unique Key	Created Date	hour
30195273	3/18/15 2:12	1/3/04 0:00
30203057	3/18/15 2:00	
30197320	3/18/15 1:58	
30194112	3/18/15 1:37	
30202379	3/18/15 1:36	
30199506	3/18/15 1:28	

Extracting Hour From Timestamp

A	B	C
Unique Key	Created Date	hour
30195273	3/18/15 2:12	1/3/04 0:00
30203057	3/18/15 2:00	
30197320	3/18/15 1:58	
30194112	3/18/15 1:37	
30202379	3/18/15 1:36	
30199506	3/18/15 1:28	

Extracting Hour From Timestamp



The screenshot shows a Microsoft Excel interface. The Home tab is selected in the ribbon. The Number dropdown menu is open, with the 'Number' option highlighted by a red box. To the right of the ribbon, the 'Format Cells' dialog box is displayed, specifically the 'Number' tab. In the preview area of the dialog box, the value '12' is shown, and 'Time' is selected from the format dropdown. Below the ribbon, there is a table with three columns: Unique Key, Created Date, and hour. The 'hour' column contains values like 2.00, 2:00, etc., corresponding to the timestamp values in column B.

A	B	C
Unique Key	Created Date	hour
30195273	3/18/15 2:12	2.00
30203057	3/18/15 2:00	2:00
30197320	3/18/15 1:58	1:58
30194112	3/18/15 1:37	1:37
30202379	3/18/15 1:36	1:36
30199506	3/18/15 1:28	1:28

Extracting Hour From Timestamp

The screenshot shows a Microsoft Excel interface with a table containing three columns: Unique Key, Created Date, and hour. The 'hour' column has a formula applied to the first cell (C2). The formula bar at the top shows the formula =HOUR(B2). The 'Number' section of the 'Format Cells' dialog box is open, with the 'Decimal places' dropdown set to 0, which is highlighted with a red box.

A	B	C
Unique Key	Created Date	hour
30195273	3/18/15 2:12	2
30203057	3/18/15 2:00	
30197320	3/18/15 1:58	
30194112	3/18/15 1:37	
30202379	3/18/15 1:36	
30199506	3/18/15 1:28	

Extracting Hour From Timestamp

The screenshot shows a Microsoft Excel interface with a table containing three columns: Unique Key, Created Date, and hour. The 'hour' column has a formula applied to the first cell (C2). A blue callout bubble points to the bottom-right corner of the cell C2, indicating that double-clicking there will expand the formula down the column. The formula bar at the top shows the formula =HOUR(B2).

A	B	C
Unique Key	Created Date	hour
30195273	3/18/15 2:12	2
30203057	3/18/15 2:00	
30197320	3/18/15 1:58	
30194112	3/18/15 1:37	
30202379	3/18/15 1:36	
30199506	3/18/15 1:28	

Double-click on lower right hand corner to expand formula to bottom of column

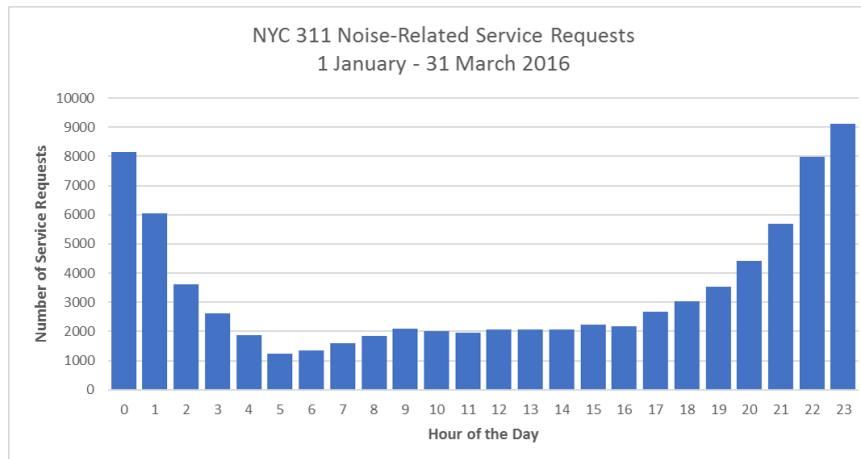
Extracting Hour From Timestamp

A	B	C
Unique Key	Created Date	hour
30195273	3/18/15 2:00	2
30203057	3/18/15 2:00	2
30197320	3/18/15 1:58	1
30194112	3/18/15 1:37	1
30202379	3/18/15 1:36	1
30199506	3/18/15 1:28	1

5. Visualizing Data

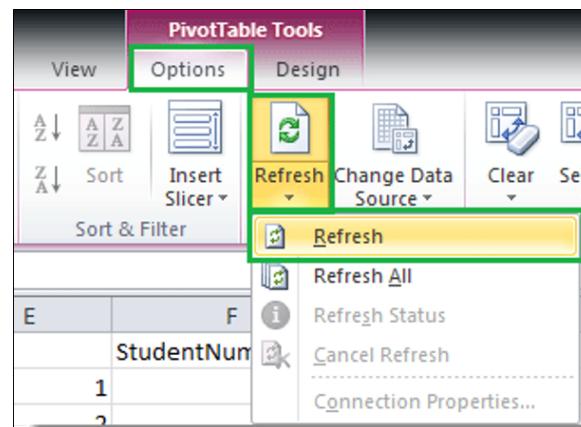
- Quickly communicate information
- Tell a clearer story
- A picture is worth a thousands words

When Are Noise Complaints Received?



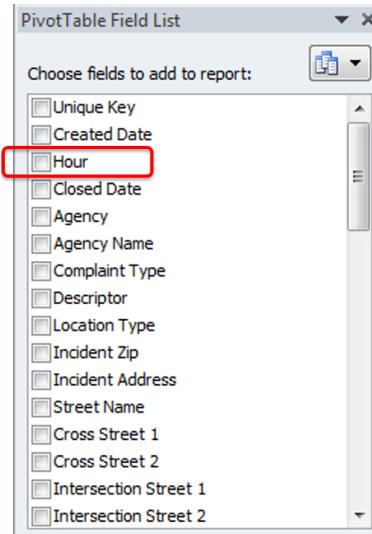
Charting Noise Complaints by Hour

- Refresh PivotTable



Charting Noise Complaints by Hour

- Refresh PivotTable
- Find `Hour` and add it to "Rows"



Charting Noise Complaints by Hour

- Refresh PivotTable
- Find `Hour` and add it to "Rows"
- Style the chart

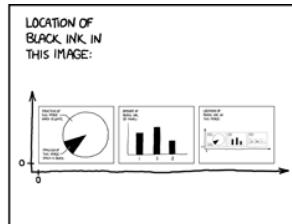
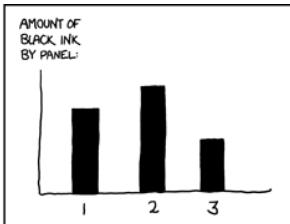
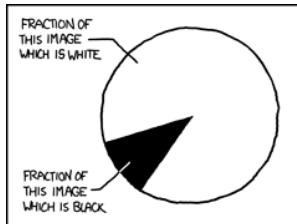
Exploratory Data Analysis

- Goal -> Discover patterns in the data
- 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 the Data

15 MIN BREAK



Source

Let's get back to our Process Map

Outline the Process

- What are the steps to creating our outputs?
- What is the best order of steps?
- How granular do we need to break this down for clarity?

Collection
Verification
Description
Sensemaking
Communication/Visualization

4 Concerns

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*

When HIPAA Gets In The Way of Health Care



Image Credit: Hipaa Nurse Shredding Papers, by [Atlantic Training](#), CC BY-SA 3.0

[Learn more](#)

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 office and ultimately to the organization as a whole*

Benefits

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

What could be some of the benefits we'd realize from what we've mapped?

What did you notice about this process?

- Knowing the problem and sequencing the steps can be harder than working with data
- Having these answers can make the analysis much easier
- This is a process you can do with any challenge, no matter how big
- Everyone has something to add, no matter how technical (or non-technical) they are

Let's Work Our Plan

NYC Community Districts

Links to Other Open Data Portals

- New York State - <https://data.ny.gov>
- US Federal Government - <https://www.data.gov/>



BUILDING A DATA-DRIVEN CULTURE



Data Driven Culture

Data Driven Culture

<http://localhost:8895/#1>

Page 81 of 122

Data Driven Culture

<http://localhost:8895/#1>

Page 82 of 122

Data Driven Culture

<http://localhost:8895/#1>

Page 83 of 122

Data Driven Culture

<http://localhost:8895/#1>

Page 84 of 122

Data Driven Culture

<http://localhost:8895/#1>

Page 85 of 122

Data Driven Culture

<http://localhost:8895/#1>

Page 86 of 122

Data Driven Culture

<http://localhost:8895/#1>

Page 87 of 122

Data Driven Culture

<http://localhost:8895/#1>

Page 88 of 122

Data Driven Culture

<http://localhost:8895/#1>

Page 89 of 122

Data Driven Culture

<http://localhost:8895/#1>

Page 90 of 122

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

[Source](#)

WRAP-UP

What We've Covered

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

What might you do different with this information when you go back to your offices?

Final Thoughts

- Data can tell a story, but doesn't speak for itself
- Analysis is the search for understanding and where we learn to tell that story
- Be good to your data and it will be good to you

Technical Support

- [Microsoft Office Support](#) - Documentation on various MS Office products
- [Data Science Cheatsheet](#) - Includes various terms and concepts related to data science
- [Open Data Handbook](#) - Guides, case studies and resources for government & civil society on the "what, why & how" of open data
- [Copy of today's handout](#)

Resources

- Harvard Government Performance Lab Results-Driven Contracting
- Carl Anderson *Creating a Data-Driven Organization*
- DJ Patil & Hilary Mason *Data Driven: Creating a Data Culture*
- IDEO Design Kit - Resource for design thinking techniques
- Datapolitan training classes

Contact Information

Richard Dunks

- Email: richard[at]datapolitan[dot]com
- Website: <http://www.datapolitan.com>
- Twitter: [@datapolitan](https://twitter.com/datapolitan)

THANK YOU!

Types of Analysis

THE NOLALYTICS
CIVIC ANALYTICS TYPOLOGY

A TYPOLOGY OF PROJECTS FORMS OUR TEMPLATE FOR DEVELOPING NEW USE CASES WITH DEPARTMENTS

See more at <http://datadriven.nola.gov/>

Civic Analytics Network

Based on the work of the [City of New Orleans, Office of Performance and Accountability](#)

Graphics: Copyright © Harvard University Ash Center (Used with Permission)

Finding a Needle in Haystack

- **Challenge:** Targets are difficult to identify or locate within a broader population
- **Opportunity:** Data analysis and predictive modeling to identify targets based on existing data



Finding the
needle in a
haystack

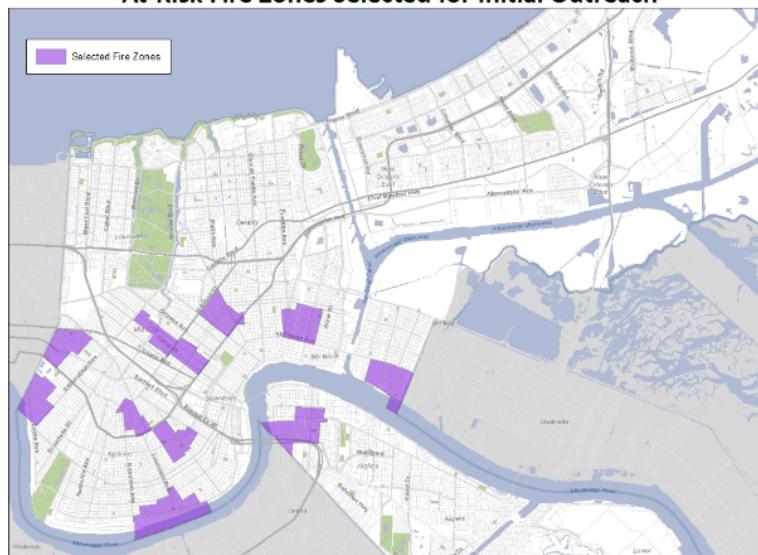
New Orleans Distributes Smoke Alarms

<http://localhost:8895/#1>

Page 103 of 122

Targeted Outreach Saves Lives

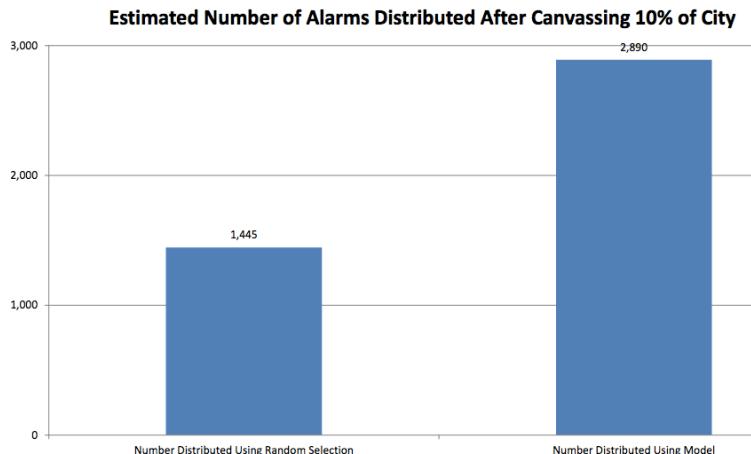
At-Risk Fire Zones Selected for Initial Outreach

<http://localhost:8895/#1>

Page 104 of 122

Targeted Outreach Saves Lives

Predictive power of model



Prioritizing Work for Impact

- Challenge: Services do not categorize high-priority cases early
- Opportunity: Data analysis and predictive modeling to prioritize cases



NYC Restaurant Inspectors Save Time



Image Credit: Grease Trap Cover Asphalt by Christopher Sessums, CC BY 2.0

Source: http://www.nyc.gov/html/dep/html/press_releases/12-71pr.shtml#.W6v2nRNKhZo

Early Warning Tools

- **Challenge:** Resources overly focused on reactive services
- **Opportunity:** Developing tools to predict need based on historic patterns



Using Data to Fight Fires

<http://localhost:8895/#1>

Page 109 of 122

Better, Quicker Decisions

- **Challenge:** Repeated decisions are made without access to all relevant information
- **Opportunity:** Developing recommendation tools for operational decisions



Cincinnati Targets Urban Blight



Image Credit: Wholtone, via [Wikimedia Commons](#)

[Learn more](#)

Optimizing Resource Allocation

- **Challenge:** Assets are scheduled or deployed without input of latest service data
- **Opportunity:** Use data to drive decisions on the deployment of resources



Shortening Public School Bus Rides



Image Credit: [Patrick Hudepohl](#), used under [Creative Commons BY-NC-SA 2.0 license](#).

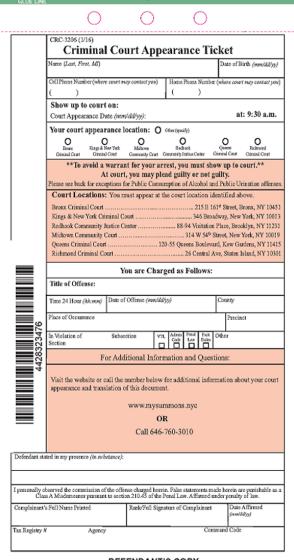
Experimenting for What Works

- **Challenge:** Services have not been assessed for impact
- **Opportunity:** Experimental testing and improvement of service options



Experimenting
for what works

Redesign of NYC Summons

	<p>CRC-3206(5/15)</p> <p>Criminal Court Appearance Ticket</p> <p>Name (Last, First, MI) _____ Date of Birth (month/day/year) _____</p> <p>Offense Number (where court issued first) _____ State/Place Number (where court issued first) _____</p> <p>() _____ () _____</p> <p>Show up to court on: _____ at: 9:30 a.m.</p> <p>Court Appearance Date (month/day/year): _____</p> <p>Your court appearance location: _____</p> <p><input checked="" type="radio"/> Bronx <input type="radio"/> Brooklyn <input type="radio"/> Manhattan <input type="radio"/> Queens <input type="radio"/> Richmond <input type="radio"/> Staten Island</p> <p><input type="checkbox"/> Read a New York Criminal Court Appearance Ticket <input type="checkbox"/> Read a New York Criminal Court Appearance Ticket <input type="checkbox"/> Read a New York Criminal Court Appearance Ticket</p> <p>*To avoid a warrant for your arrest, you must show up to court.*^a</p> <p>At court, you may plead guilty or not guilty.</p> <p>Please see back for more information on how to appear in criminal court and what to do if you are arrested.</p> <p>Court Locations: You must appear at the court location identified above.</p> <p>Bronx Criminal Court - 215 E. 167th Street, Bronx, NY 10451 Kings & New York Central Court - 345 Broadway, New York, NY 10013 Brooklyn Criminal Court - 125 1/2 Fulton Street, Brooklyn, NY 11201 Midtown Community Court - 314 W 49th Street, New York, NY 10019 Queens Criminal Court - 220-55 Queens Boulevard, Kew Gardens, NY 11415 Richmond Criminal Court - 24 Centre Ave, Staten Island, NY 10304</p> <p>You are Charged as Follows:</p> <p>Title of Offense: _____ Date of Offense (month/day) _____ County _____</p> <p>Place of Occurrence: _____ District _____</p> <p>In Violation of: _____ Section: _____ Part: _____ Subpart: _____ Date: _____</p> <p><input checked="" type="checkbox"/> Bronx <input type="checkbox"/> Brooklyn <input type="checkbox"/> Manhattan <input type="checkbox"/> Queens <input type="checkbox"/> Richmond <input type="checkbox"/> Staten Island</p> <p>For Additional Information and Questions:</p> <p>www.nycsummons.ny.gov OR Call 646-760-3010</p> <p>Defendant stated in my presence (in substance): _____</p> <p>I personally observed the commission of the offense charged herein. False statements made herein are punishable as a Class A Misdemeanor pursuant to section 210.05 of the Penal Law. A Plaintiff under penalty of law.</p> <p>Complainant's Name Printed: _____ Rank/Title/Signature of Complainant: _____ Date Attested (month/day/year): _____</p> <p>Agency: _____ Tax Registry # _____ Command Code: _____</p> <p>The person described above is summoned to appear at NYC Criminal Court, _____ located at: _____</p> <p>Date of Appearance (month/day): _____ At 9:30 a.m.</p> <p>DEFENDANT'S COPY</p>
-----------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Definition of Open Data

Open data is data that can be freely used, shared and built-on by anyone, anywhere, for any purpose

- Open Knowledge International

Key Features of Open Data

- Availability and access
- Reuse and redistribution
- Universal participation

Open Data Benefits

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

Keeping NYC Accountable on Parking Tickets



Image Credit: Parking Violator by Atomische * Tom Giebel, CC BY-NC-ND 2.0

Source: <http://iqrantny.tumblr.com/post/87573867759/success-how-nyc-open-data-and-reddit-saved-new>

Open Data Concerns

- Privacy ([PII](#) and [the Mosaic Effect](#))
- Accuracy
- Security

When Good Data Goes Bad



Image Credit: Kenny Louie, CC BY 2.0, via Wikimedia Commons.jpg)

[Gawker matches Taxi and Limousine Data with Paparazzi Photos](#)