



Management Academy
Analytics for the Data-Driven Manager
Instructor: Richard Dunks
Course Code: C4311

13-14 January 2016

Data**politan**

Data Solutions for the Modern Metropolis

WELCOME

INTRODUCTIONS

Name _____
Office _____
One thing you hope to get out of today's class _____

Goals for the Course

- Discuss the data-driven decision making process as it relates to city government
- 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 using data in the decision-making process
- Better understand how to build a data-driven culture
- 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

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

Why Do We Collect Data?

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

Why Do We Publish Laws?



WHAT GOOD ARE LAWS IF WE DON'T KNOW HOW THEY'RE IMPLEMENTED?

That's what data tells us

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

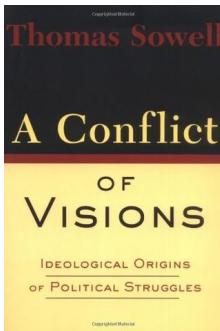
Success: How NYC Open Data and Reddit Saved New Yorkers Over \$55,000 a Year

Before Open Data:  After Open Data: 

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

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

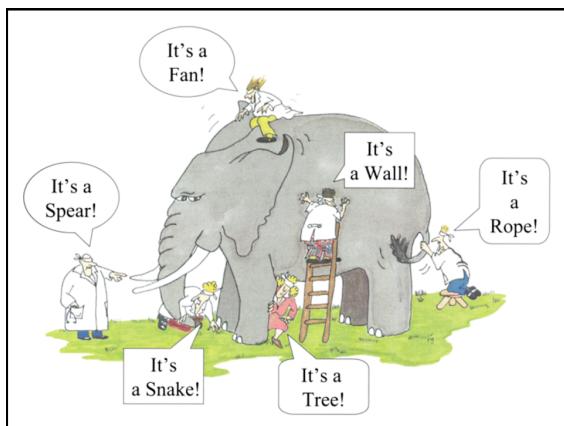
— Thomas Sowell,
A Conflict of Visions

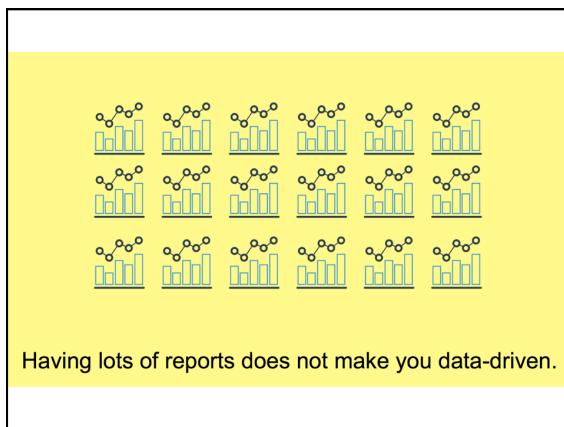


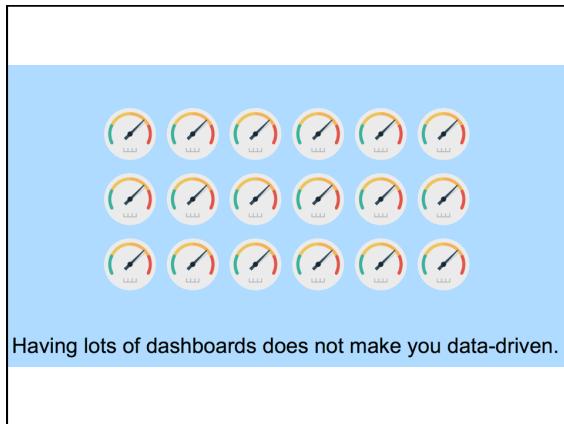
WHAT DOES “DATA-DRIVEN” MEAN IN THE CONTEXT OF CITY GOVERNMENT?

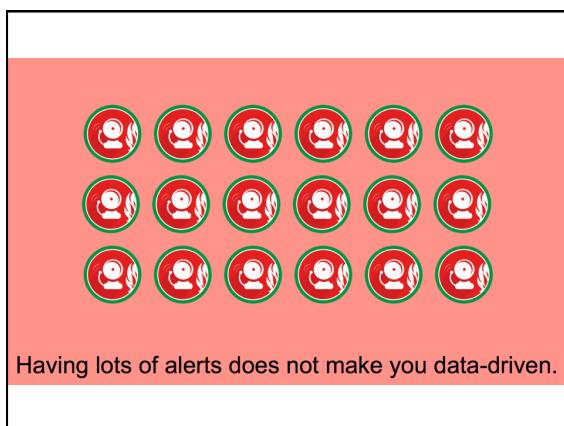
Discuss in your group

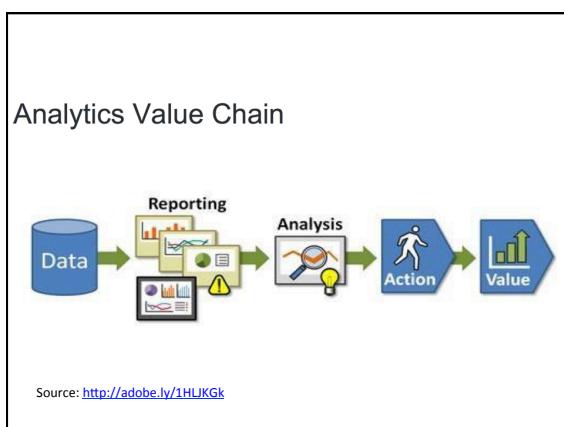








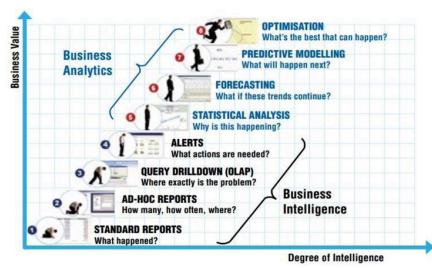




TYPES OF ANALYSIS

**DATA ANALYTICS?
BUSINESS ANALYTICS?
DATA SCIENCE?**

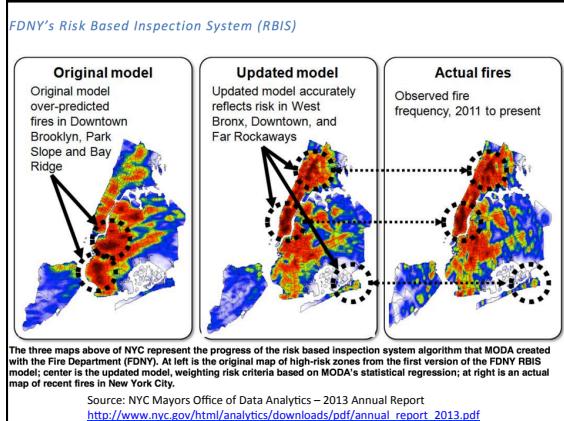
Levels of Analytics



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

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

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.

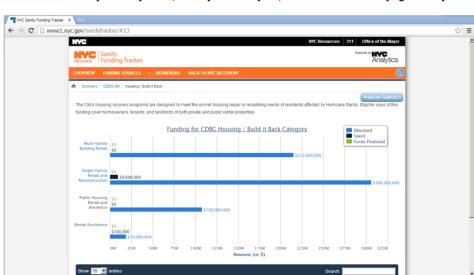
Source: <http://www.charlotteobserver.com/news/business/article32617293.html>

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

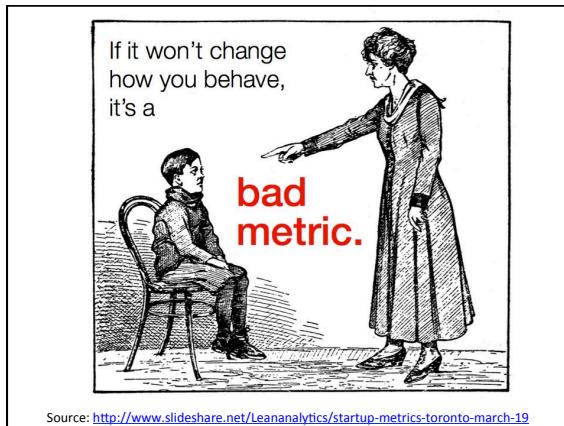
Tracking information on Sandy recovery funds, built by NYC analytics, is available at www1.nyc.gov/sandytracker



The chart shows the following approximate data:

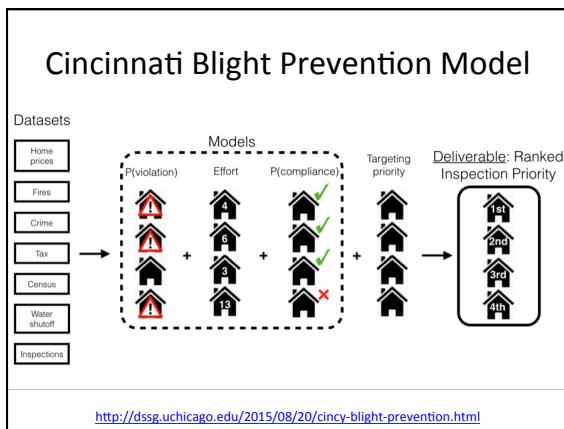
Category	Allocated Funds	Funds Received
Multi Family Housing	\$1,100,000,000	\$1,100,000,000
Single Family Housing	\$100,000,000	\$100,000,000
Residential Businesses	\$100,000,000	\$100,000,000
Public Housing	\$100,000,000	\$100,000,000
Nonprofit Organizations	\$100,000,000	\$100,000,000

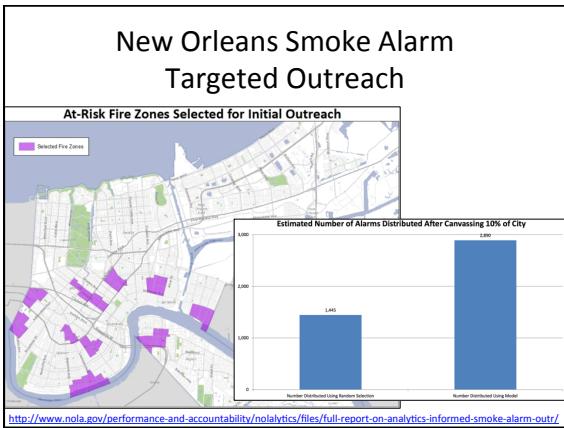
Source: NYC Mayors Office of Data Analytics – 2013 Annual Report
http://www.nyc.gov/html/analytics/downloads/pdf/annual_report_2013.pdf



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





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

WHAT KIND OF ANALYSIS DOES YOUR OFFICE DO?

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*

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 the university and ultimately to the CUNY system 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/inbloom-shuts-down-amid-privacy-fears-over-student-data-tracking>

Benefits

- Time, money, lives saved



FOR IMMEDIATE RELEASE
October 18, 2012
No. 71

NEW YORK CITY BUSINESS INTEGRITY COMMISSION, DEPARTMENT OF ENVIRONMENTAL PROTECTION, AND MAYOR'S OFFICE OF POLICY AND STRATEGIC PLANNING LAUNCH COMPREHENSIVE STRATEGY TO HELP BUSINESSES COMPLY WITH GREASE DISPOSAL REGULATIONS

*Enforcement Effort Will Target Areas With Highest Concentration of Yellow Grease Production;
DEP Launches Educational Video for Restaurant Industry on Proper Grease Disposal*

http://www.nyc.gov/html/bic/downloads/pdf/pr/nyc_bic_dep_mayoroff_policy_10_18_12.pdf

Benefits

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

Being data-driven doesn't mean



blindly following data.

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

Being data-driven means having...



an open, sharing culture

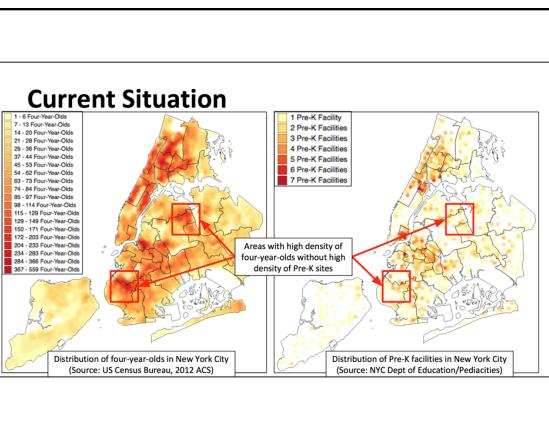
No data hoarding or silos. Bring data together to
create rich contexts. Connect the dots.

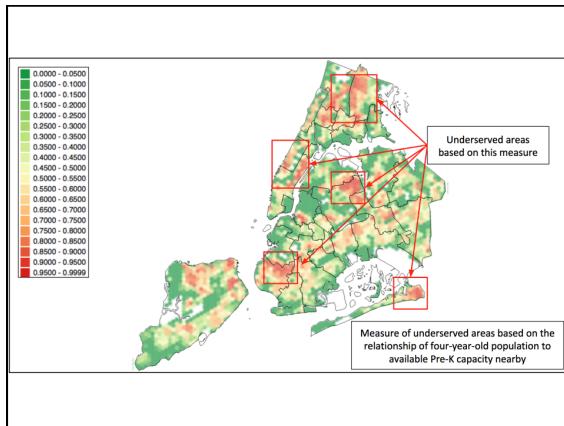
WHAT CONCERNS DO YOU HAVE WITH RESPECT TO ANALYTICS IN YOUR JOB?

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

Group Exercise – Universal Pre-K

- Define the analytical problem 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

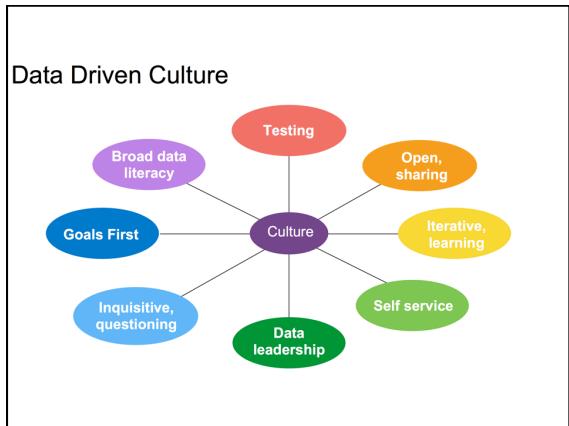




Goals for this Afternoon

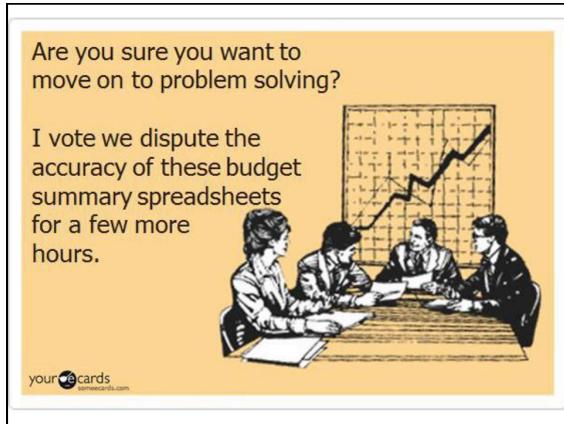
- Discuss the features of a data-driven culture
- Demonstrate the value of well-designed information visualizations
- Familiarize you with exploratory data analysis and question-driven analysis
- Practice communicating analytical findings
- Become familiar with key features and issues with government open data

BUILDING A DATA-DRIVEN CULTURE









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.

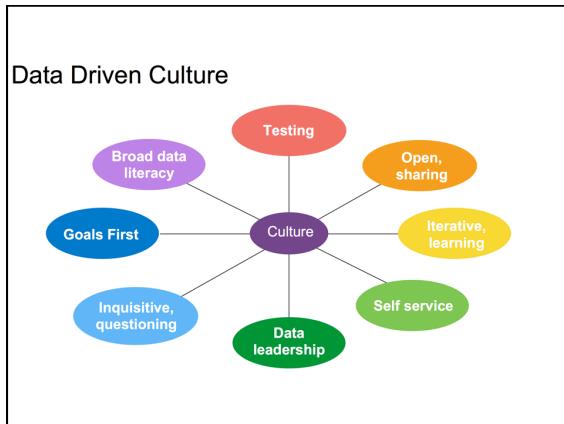
Change should not just be top-down

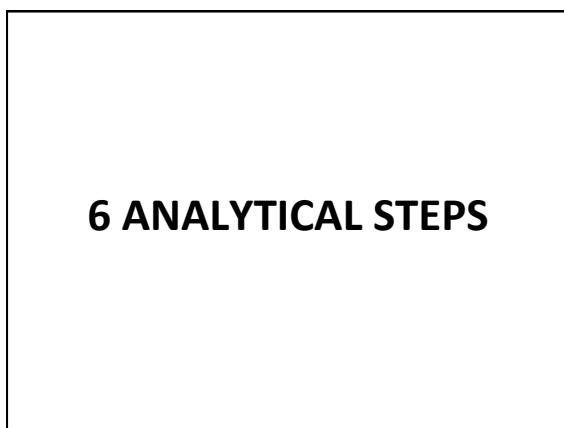


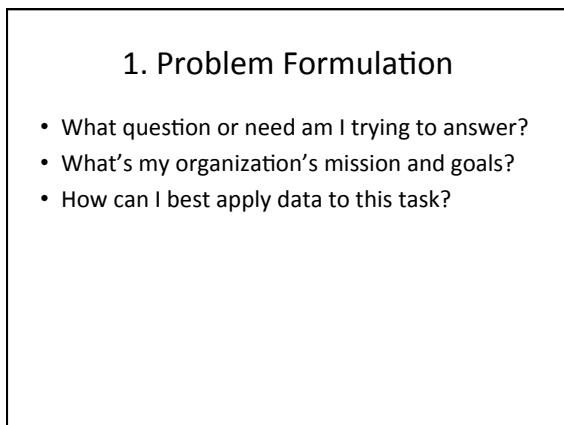
but bottom up too

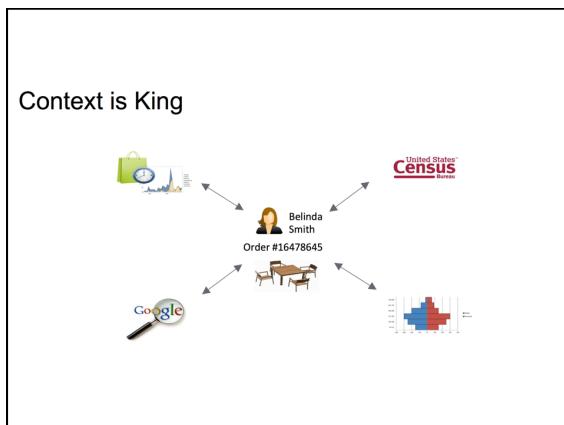
Everyone in org has role and responsibility through "leveling up" their data skills, mutual mentoring, and embedding data into their processes.

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

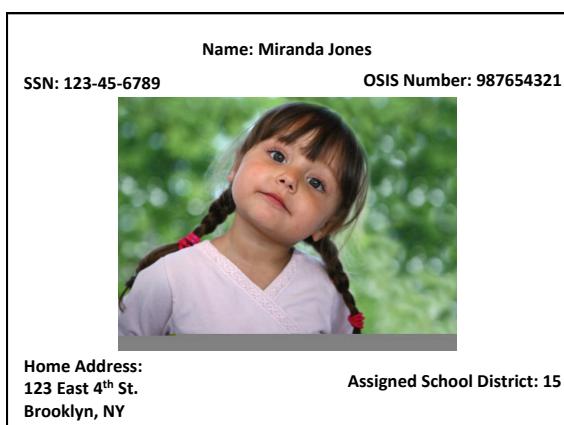












2. Data Gathering/Preliminary Analysis

- 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

The New York Times

TECHNOLOGY

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

By STEVE Lohr AUG. 18, 2014



Monica Rogati, Jawbone's vice president for data science, with Brian Will, 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



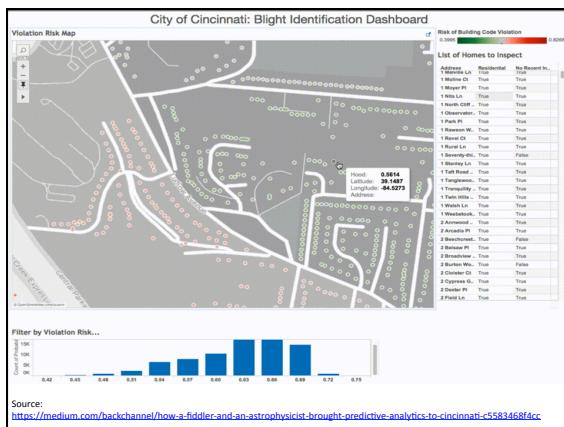
<http://www.businessinsider.com.au/excel-partly-to-blame-for-trading-loss-2013-2>

5. Verification – London Whale

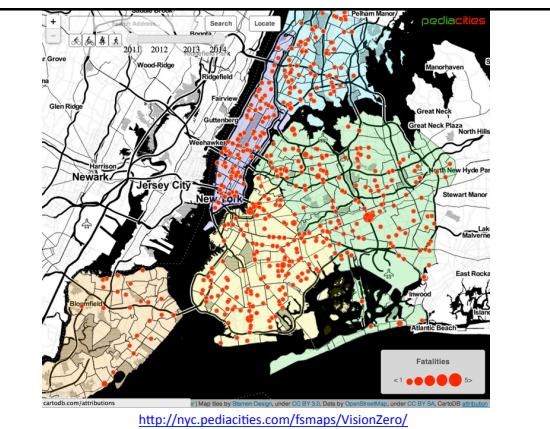
- Caused largely by Excel mistakes:
 - Manual data errors
 - Manual copy and paste
 - Simple formula error that hid volatility
- Fined over \$1 billion for poor internal oversight of trading activities

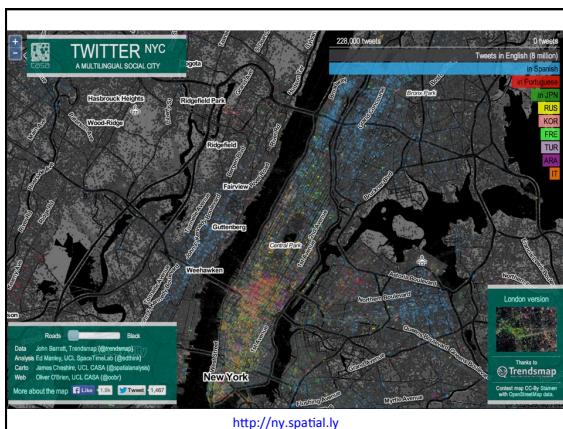
6. Visualization

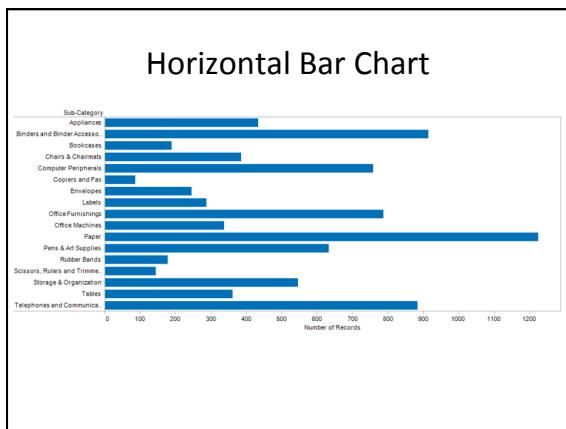
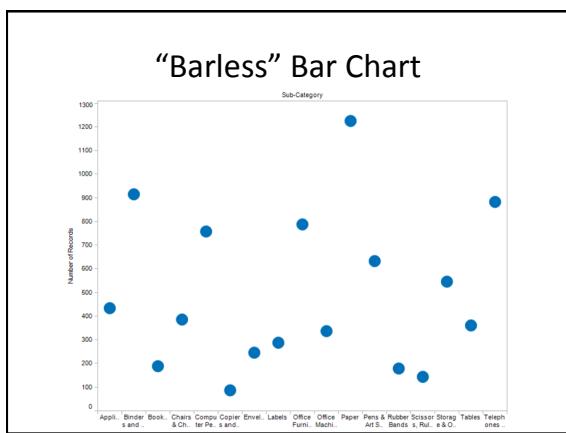
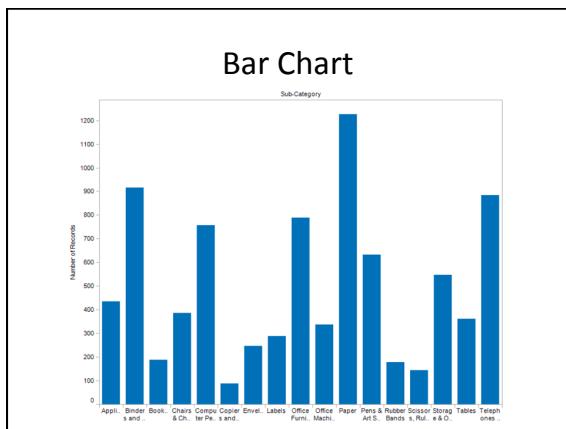
- “A picture is worth a thousand words”
- Communicate results clearly and concisely
- Help to better understand your data
- The eyes have a much higher bandwidth into the brain

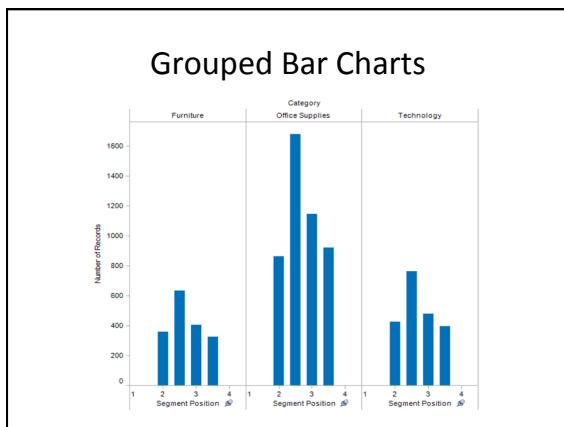
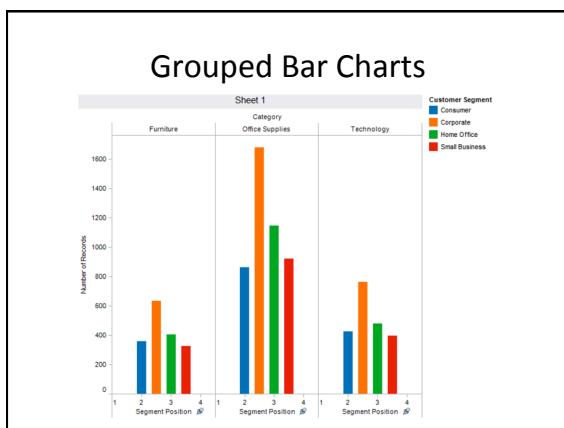
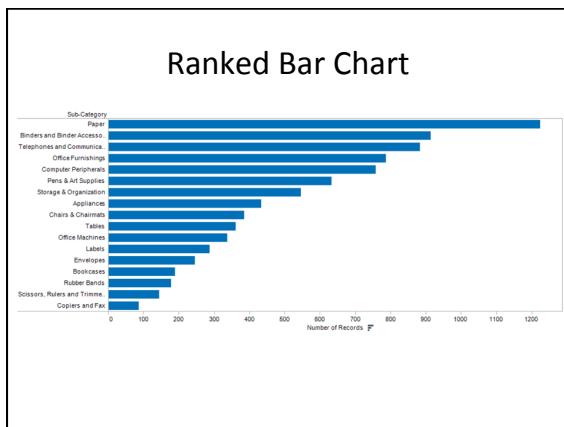


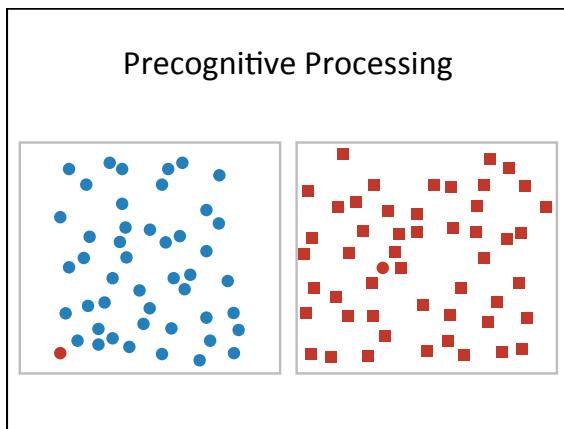
HOW DOES YOUR OFFICE ANALYZE DATA?

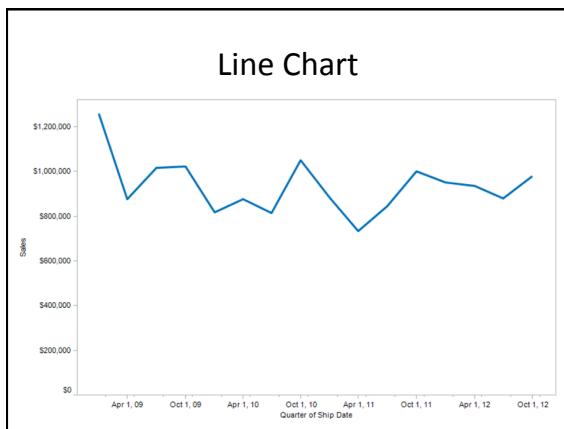


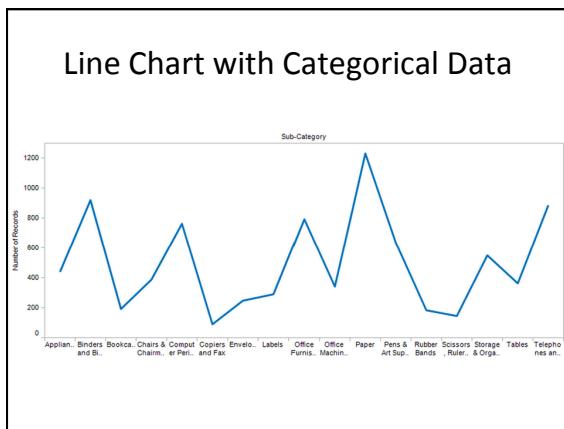


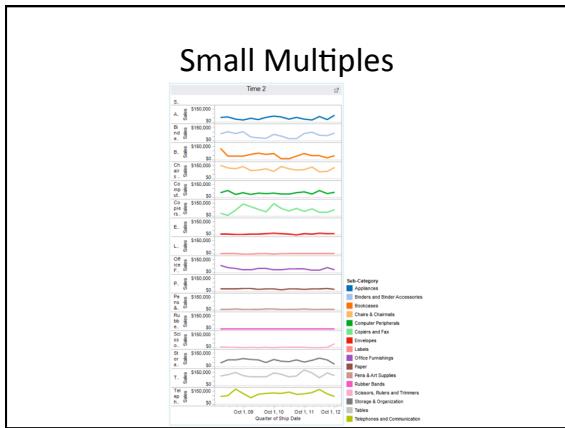
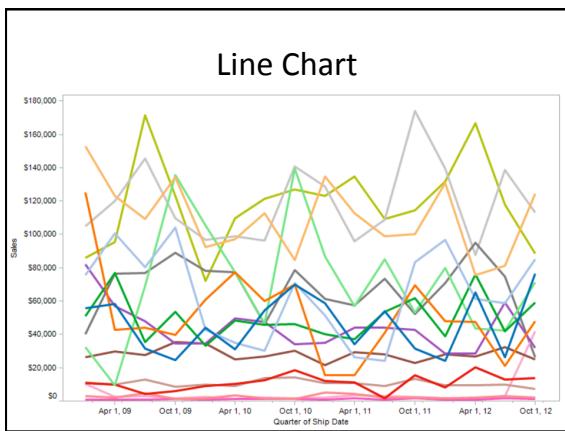
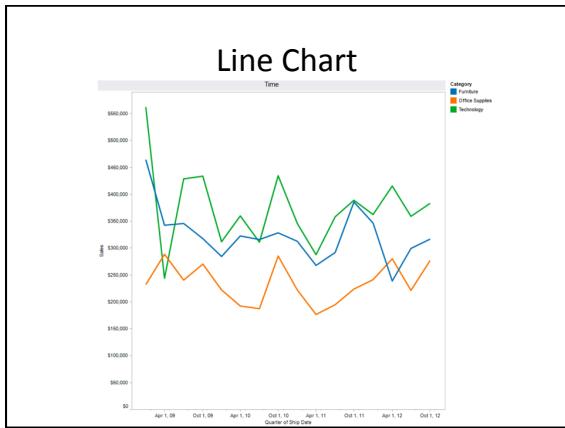


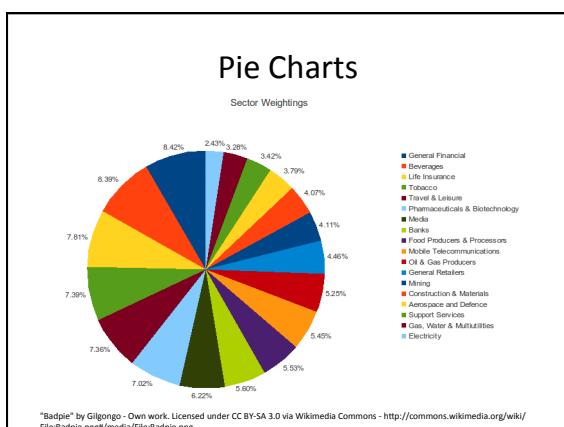
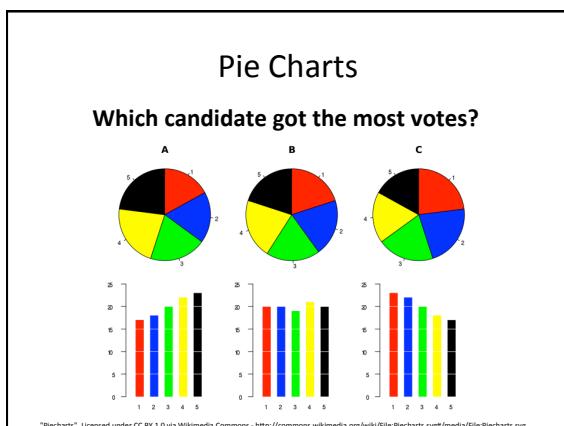
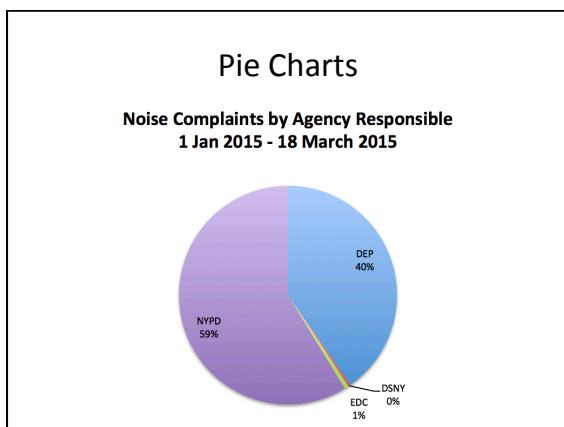




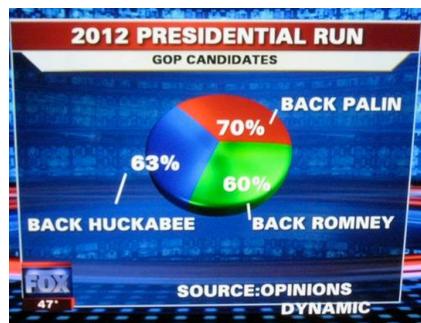






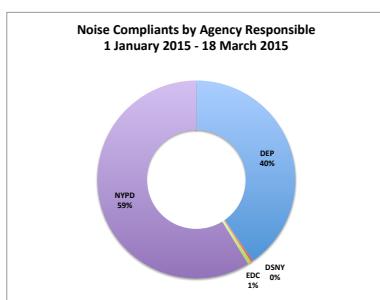


Pie Charts



<http://simplystatistics.org/2012/11/26/the-statisticians-at-fox-news-use-classic-and-novel-graphical-techniques-to-lead-with-data/>

Donut Charts



How do you learn to make good charts?
...Make a lot of bad charts

Definition of Open Data

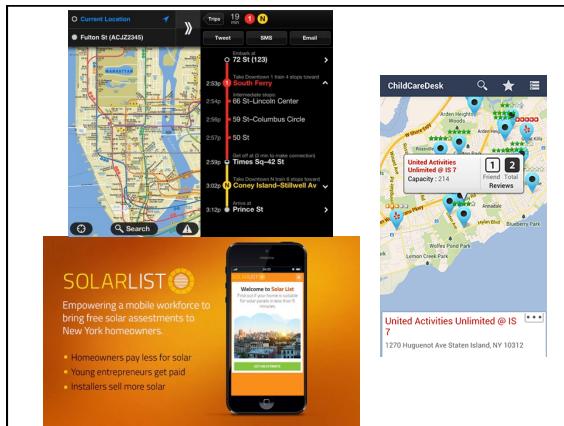
- Definition:
 - Open data is data that can be freely used, shared and built-on by anyone, anywhere, for any purpose
 - Key Features
 - Availability and access
 - Reuse and redistribution
 - Universal participation

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



Open Data Benefits

- Transparency
 - Releasing social and commercial value



Open Data Benefits

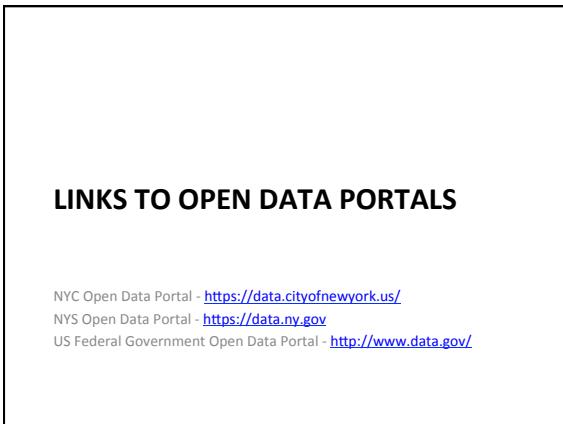
- Transparency
- Releasing social and commercial value
- Participation and engagement

 Voting Information Project helps voters find information about their elections with collaborative, open-source tools.

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

Open Data Concerns

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



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

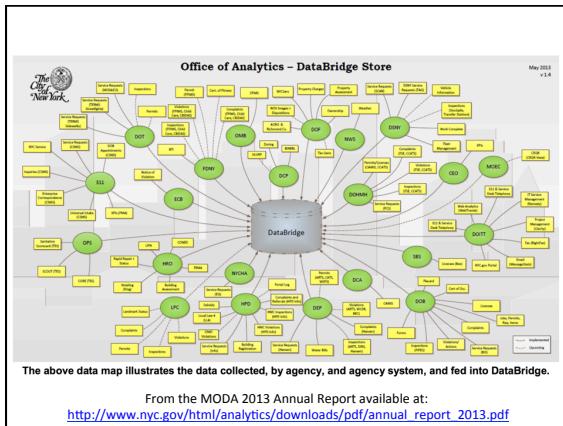


Analytical Resources

- 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

- 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
 - Key features of a data-driven culture

Goals for the Course

- Discuss the data-driven decision making process as it relates to city government
 - 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

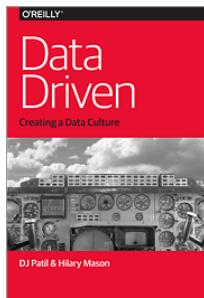
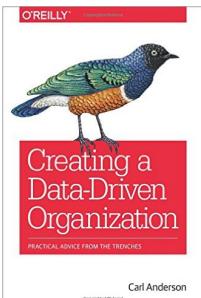
- Better understand using data in the decision-making process
- Better understand how to build a data-driven culture
- Better understand the analytics process
- Better understand the value of data, particularly open data
- Better understand the role of analysts and managers in the decision-making process



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



Contact Information

Instructor

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

THANK YOU!

REFERENCE SLIDES

(Not presented in class)

INTRODUCTION TO STATISTICS

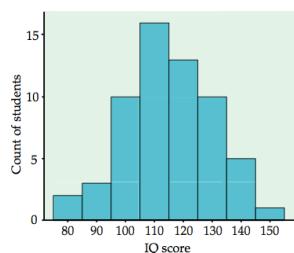
“We are drowning in information and starving for knowledge.”
-Rutherford D. Roger

Why Statistics?

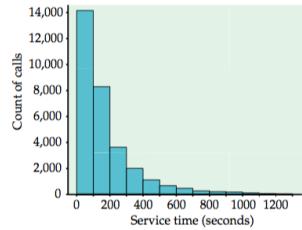
- Tools for extracting meaning from data
- Commonly understood ways of communicating meaning to others

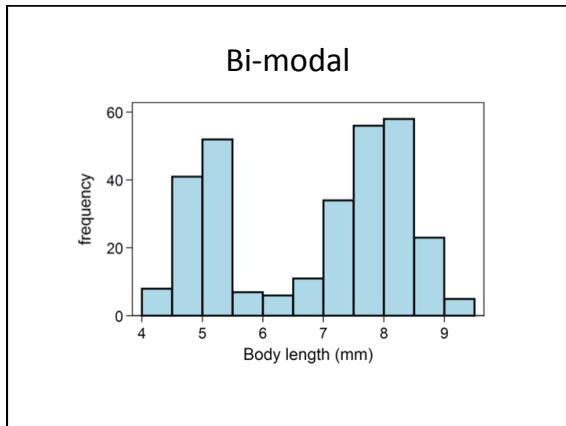
DATA DISTRIBUTIONS

Normal Distribution



Long-tail Distribution





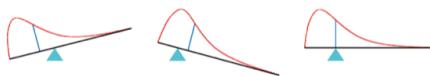
DESCRIBING DATA

Mean

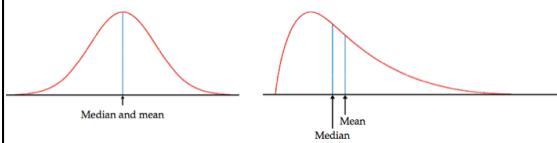
- A representative value for the data
- Usually what people mean by “average”
- Calculate by adding all the values together and dividing by the number instances
 - Example: Calculate mean height of everyone in class
- Sensitive to extremes

Median

- The “middle” value of a data set
 - Center value of a data set with an odd number of values
 - Sum of two middle values divided by 2 if the number of items in a data set is even
- Resistant to extreme values

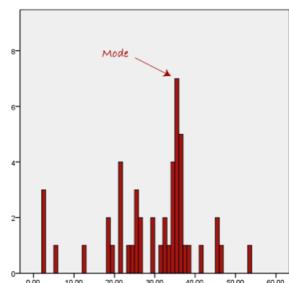


Mean vs Median



Mode

- The most frequent value in a dataset



Measures of Central Tendency

- Quantitative data tends to cluster around some central value
- Mean is a more precise measure and more often used
- Median is better when there are extreme outliers
- Mode is used when the data is categorical (as opposed to numeric)

Measures of Variability

- Describe the distribution of our data
- Help us understand how well the measures of central tendency represent the data

Range

- The gap between the minimum value and the maximum value
- Calculated by subtracting the minimum from the maximum

Quartiles

- Median splits the data set into two equal groups
- Quartiles split the data into four equal groups
 - First quartile is 0-25% of the data
 - Second quartile is 25-50% of the data
 - Third quartile is 50-75% of the data
 - Fourth quartile is 75-100% of the data

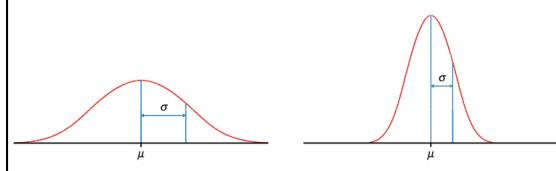
Inter-Quartile Range



<https://community.qlik.com/blog/qlikviewdesignblog/2014/08/18/recipe-for-a-box-plot>

Standard Deviation

- The average distance of each data point from the mean
- Larger the standard deviation, the greater the spread

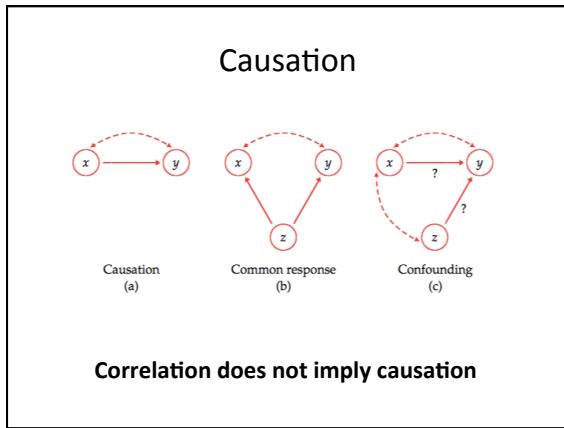
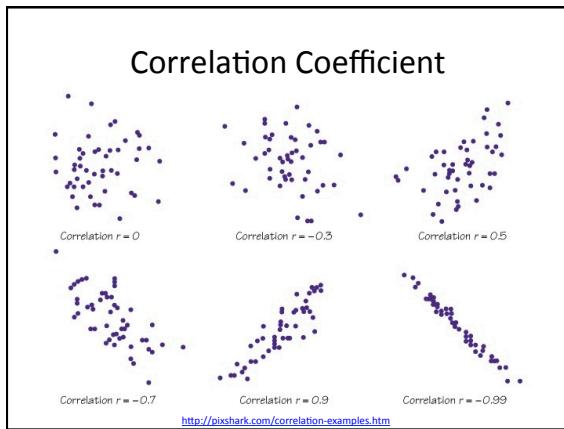


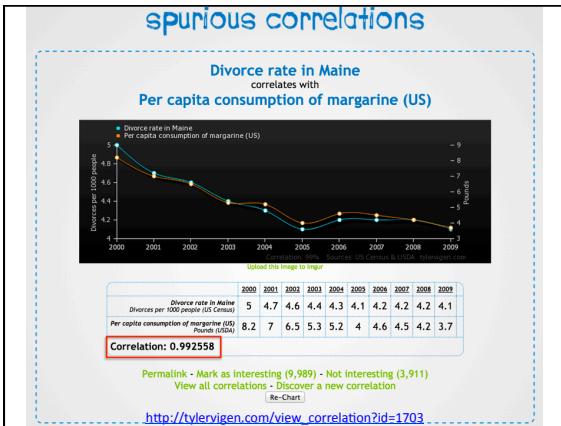
Correlations

Scatterplot of Height and Weight

A scatterplot titled "Scatterplot of Height and Weight". The x-axis is labeled "Height" and ranges from 50 to 75. The y-axis is labeled "Weight" and ranges from 0 to 160. The plot shows a positive correlation with data points clustered around a diagonal line.

How do we measure this relationship?





NYC Open Data Portal

NYC OpenData 1100+ Datasets Available

Showing All Types in category All Categories Hide Titles

Data Catalog

Search

Name	Description	Popularity	Type
WiFi Hotspot Locations	Information about WiFi hotspots in the city with basic descriptive information.	102,244 views	Geospatial
311 Service Requests from 2010 to Present	Social Services: All 311 Service Requests from 2010 to present. This information is automatically updated daily.	40,221 views	Text
Subway Entries & Exits	Transportation: mta, metropolitan transportation authority..	40,089 views	Geospatial
MTA Data	Transportation: traffic, vehicles, route, schedules, open web information pertaining to MTA Metropolitan Transportation Authority of the State of New York subways, buses, commuter rail, bridges, and tunnels	14,791 views	Geospatial
Restaurant Inspection Results	Health: NYC restaurant inspection results	25,704 views	Text

<https://nycopendata.socrata.com/data>

NYC Open Data Portal – 311 Data

NYC OpenData 1100+ Datasets Available

311 Service Requests from 2010 to Present

Manage Find in this Dataset Filter Sort & Roll-Up Add a New Filter Condition

Unique Key Created Date Closed Date

Unique Key	Created Date	Closed Date
1 20090371	10/19/2014 02:57:13 AM	
2 2009003	10/19/2014 02:51:15 AM	
3 2009227	10/19/2014 02:14:44 AM	
4 2009249	10/19/2014 02:13:30 AM	
5 2009017	10/19/2014 02:13:26 AM	
6 2009175	10/19/2014 02:09:33 AM	
7 20090900	10/19/2014 02:06:57 AM	10/19/2014
8 20094006	10/19/2014 02:06:00 AM	
9 20097010	10/19/2014 02:05:21 AM	
10 20097987	10/19/2014 02:05:20 AM	
11 20094997	10/19/2014 02:04:52 AM	
12 2009621	10/19/2014 02:04:31 AM	10/19/2014
13 20094357	10/19/2014 02:02:08 AM	10/19/2014

Filter Conditional Formatting Sort & Roll-Up Filter

Filter this dataset based on contents.

Unique Key = is = + Add a New Filter Condition

Never created a filter before? Watch a short tutorial video here.

Download 311 Data

The screenshot shows a 'Filter' interface for a dataset. It includes a 'Filter' dropdown and a 'Filter this dataset based on contents.' section. Two filters are applied:

- Created Date** is between 06/01/2014 12:00:00 AM and 09/01/2014 12:00:00 AM
- Complaint Type** is noise

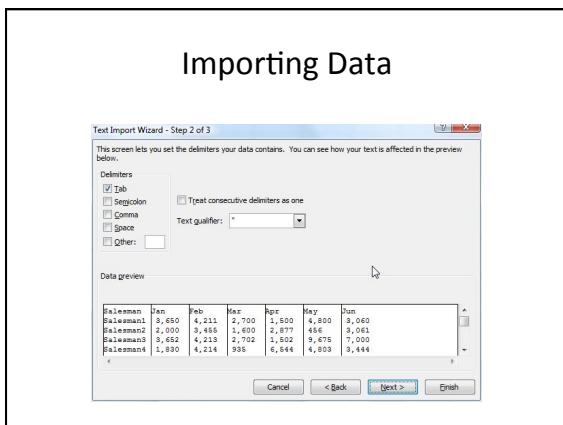
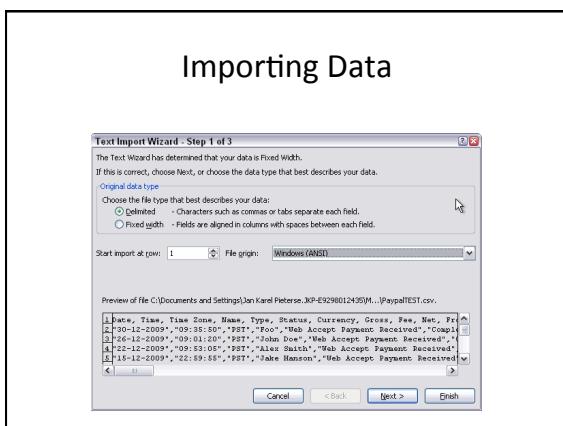
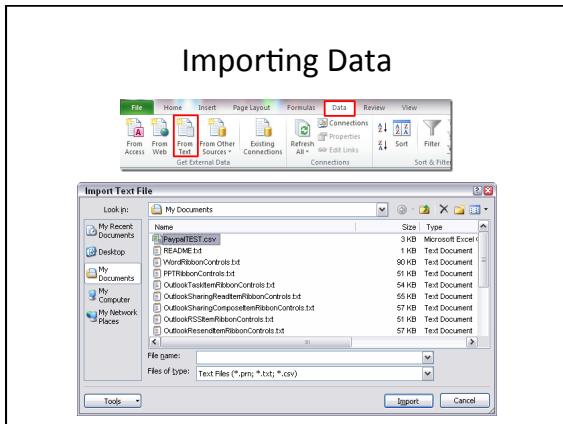
Download 311 Data

The screenshot shows a 'Download' menu. Under 'Download As', 'CSV' is highlighted with a red box.

- Export
- SODA API
- OData
- Print
- Download
- Download As
 - CSV
 - JSON
 - PDF
 - RDF
 - RSS
 - XLS
 - XLSX
 - XML

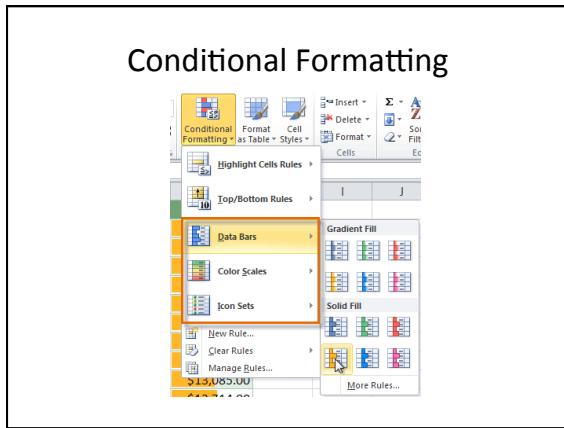
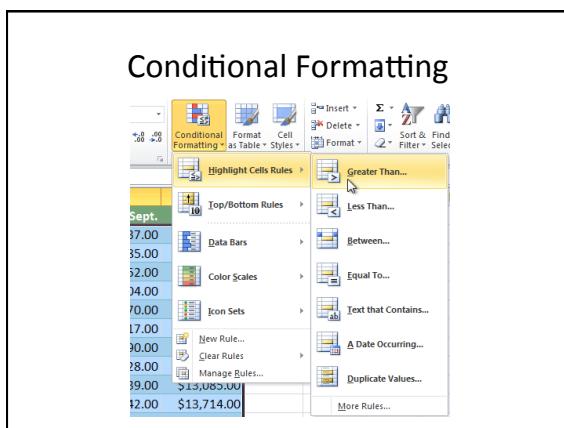
IMPORTING DATA

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



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”

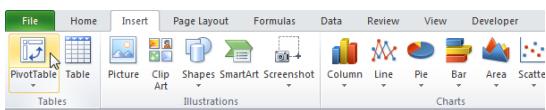


Conditional Formatting - Examples

\$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

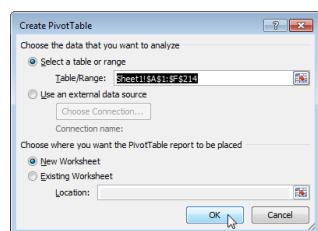
PivotTables

- 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
