The Good, the Bad, and the Ugly of Data Analytics

Albert M. Lai, PhD
Assistant Professor
Department of Biomedical Informatics
The Ohio State University

HIMSS
CENTRAL & SOUTHERN OHIO *Chapter*

HIMSS

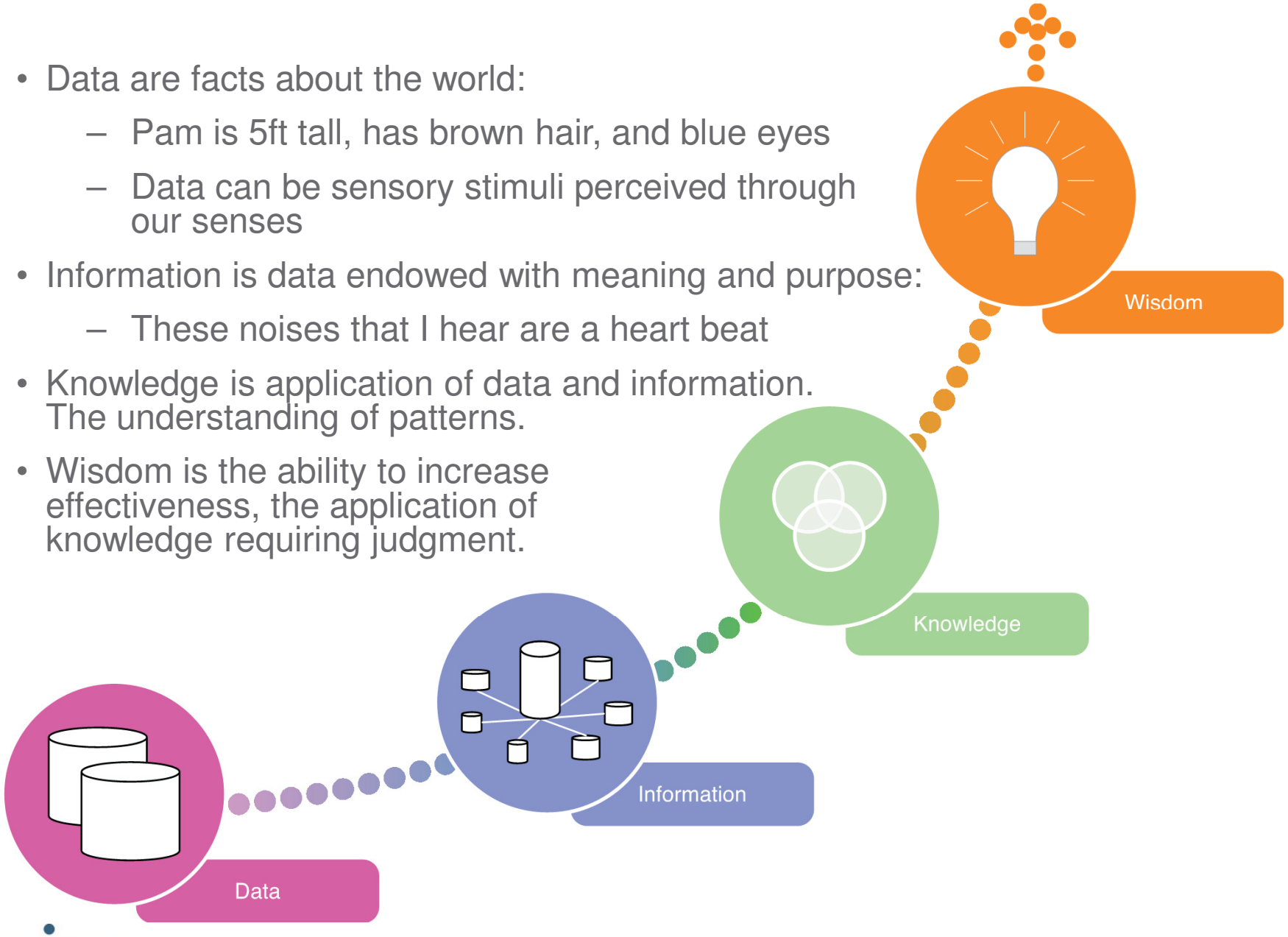
CENTRAL & SOUTHERN OHIO *Chapter*

What is Data Analytics?

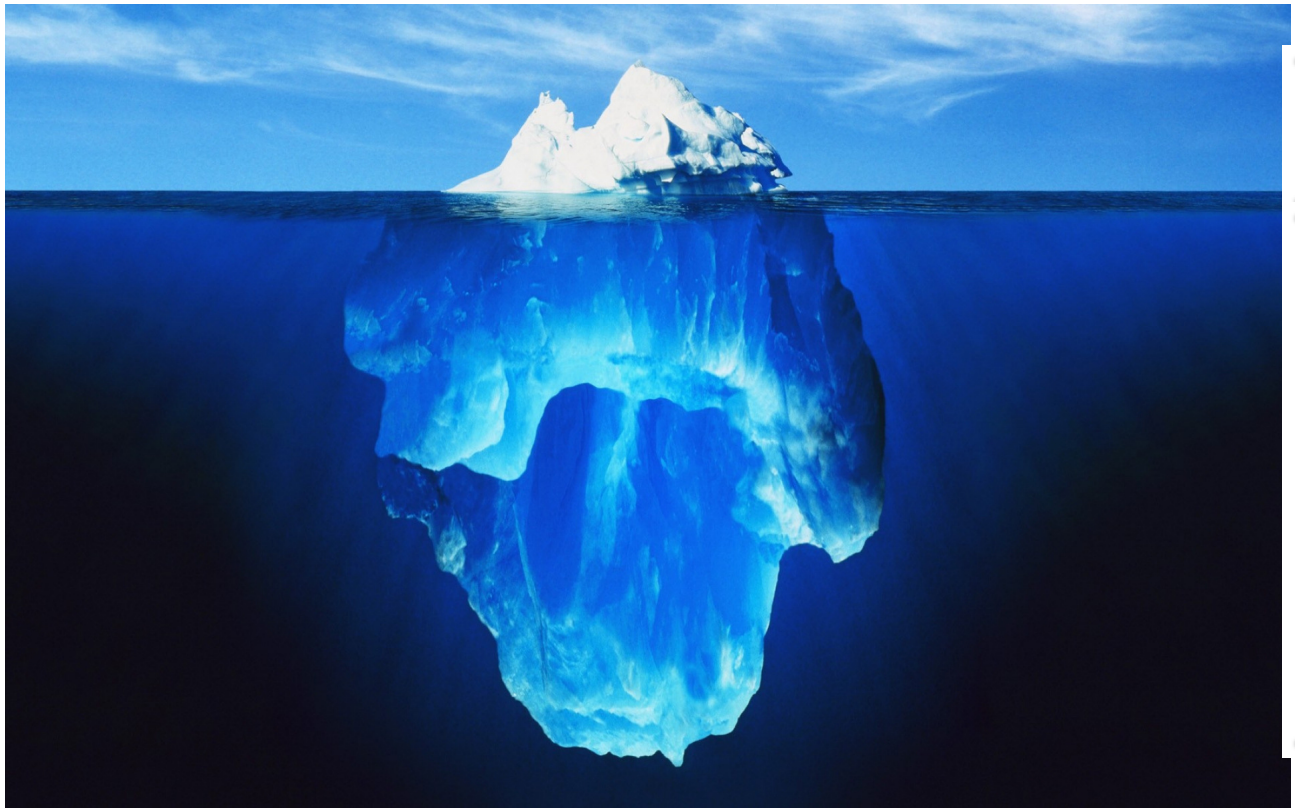
- Discovering associations and understanding patterns and trends within the data
- Science of examining raw data with the purpose of drawing conclusions about that information
- Transforming the growing amount of data into actionable information to support strategic and tactical decision making

HIMSS CENTRAL & SOUTHERN OHIO Chapter

- Data are facts about the world:
  - Pam is 5ft tall, has brown hair, and blue eyes
  - Data can be sensory stimuli perceived through our senses
- Information is data endowed with meaning and purpose:
  - These noises that I hear are a heart beat
- Knowledge is application of data and information. The understanding of patterns.
- Wisdom is the ability to increase effectiveness, the application of knowledge requiring judgment.



The Good, The Bad & The Ugly



Data analysis & presentation

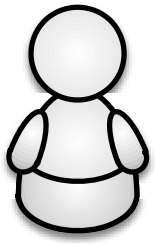
Data acquisition from clinical, financial, administrative, research systems; data integration; data cleaning; data warehousing; data governance; data provenance

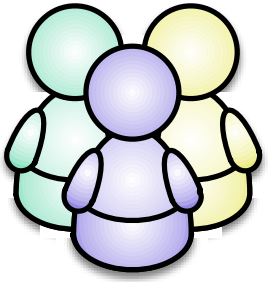
What can I do with Data Analytics? (The Good)

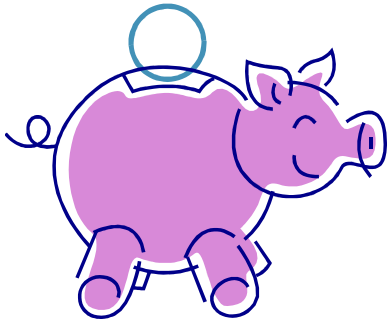
- Waste & Care Variability Reduction
- Population Health Management
- Quality Improvement
- Risk management
- Cost containment
- Predictive analytics

Why do I care?

- Not currently being done much
- Achieving or maintaining a competitive edge
- Institute for Healthcare Improvement Triple Aim:
  - Improving patient experience of care (including quality and satisfaction)
  - Improving health of populations
  - Reducing cost per capita







Common Healthcare Applications of Data Analytics

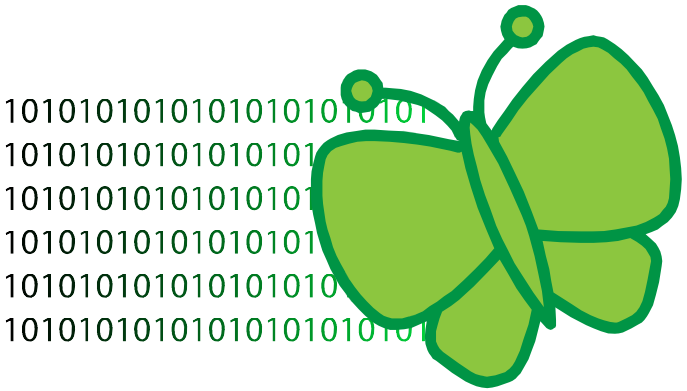
- Quality Improvement
- Syndromic Surveillance
- Understanding Readmissions
- Length of Stay Analysis
- Predictive Modeling

OSU Data Analytics Initiatives

- Predicting Heart Failure for Readmissions
- Predicting Acute Kidney Injury in the ICU
- Antimicrobial Stewardship

Process of Data Analytics

- Inspecting, cleaning, transforming, and modeling data with the goal of discovering useful information, suggesting conclusions, and supporting decision making



Transactional Databases vs. Data Warehousing

	Transactional Database	Data Warehouse
Definition	Collection of (health) data organized for storage, accessibility, and retrieval	Integrates copies of transactional data from disparate source systems for analytical use
Optimization	Performing read-write operations on individual transactions	Reading/retrieving large datasets and aggregating data
Orientation	Usually patient oriented	Subject oriented (

Transactional Databases vs. Data Warehousing (cont.)

	Transactional Database	Data Warehouse
Data Organization	Complex table structures with many joins. Normalized data (no duplicate data)	Organized to facilitate reporting & analysis, not quick transactions. Generally much fewer tables & simpler structure
SLA	99.99% uptime	Flexible
Update Frequency	Real-time	Batched ETL (e.g. 24hrs, weekly)

Finding and Cleaning EHR Data for use in Analytics (The Bad)

- Data may be in the EHR, but it's not in a readily usable format
- Just because it's been scanned into the system doesn't make it accessible
- Data frequently in PDF files
- Just because it was typed into the system doesn’t mean you can easily use it or even find it
- Even if it's in a coded/structured field, finding it is still challenging

Finding a champion

- Hospital leadership tends to be excited about the opportunities for data analytics
- Individual stakeholders tend not to be
  - Most of the models generated are not proven for effectiveness
  - Need to find a clinical (or boots on the ground) champion to support the process

Challenges with Secondary Use

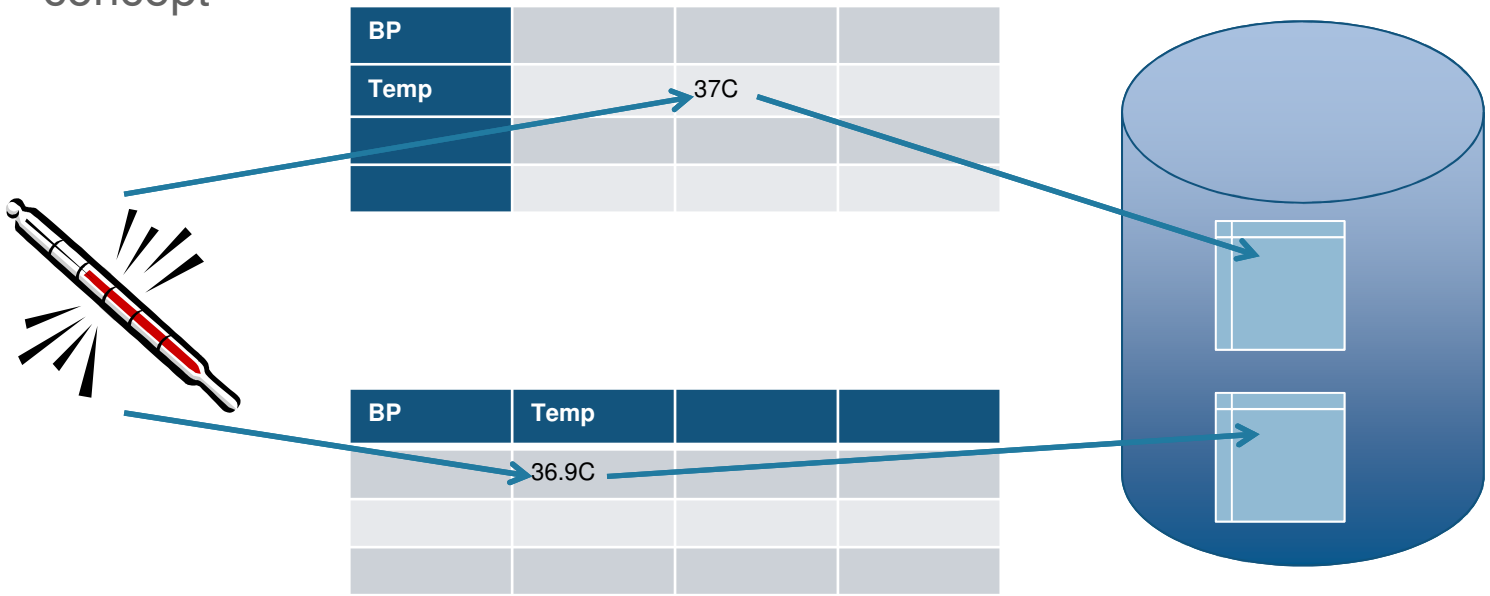
- Will need to clean the data
- Challenging to find the right data rather than using what is easy to obtain
- Data is currently truly collected for billing purposes, not clinical care
- Structured vs. Unstructured Data
  - Sometimes the data will be structured and sometimes the data will be unstructured
  - Sometimes data that you'd expect to be structured, isn't (e.g., ejection fractions)

Defining Data Provenance

- There will be an increased need to define data provenance
- Just because data is available from one source doesn't mean that it's the right place to get it from
  - E.g., For people with best of breed systems, how do you actually calculate your census for a given day?
  - Should have a consistent way of doing this for every analysis

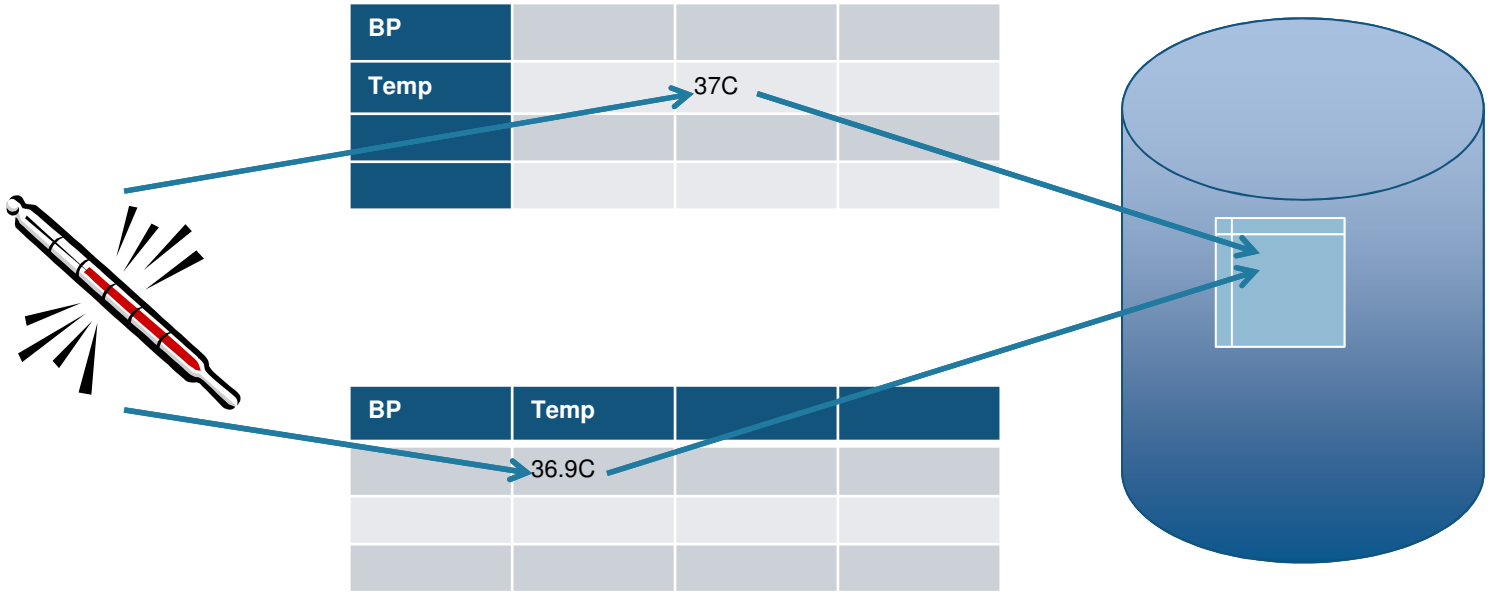
Forms

- Data can be input into many locations in an EHR
- Data from two different forms may not go into the same place on the backend database, even though they may contain the same data concept



Forms (cont.)

- You really want all data regarding the same thing to go into the same place.

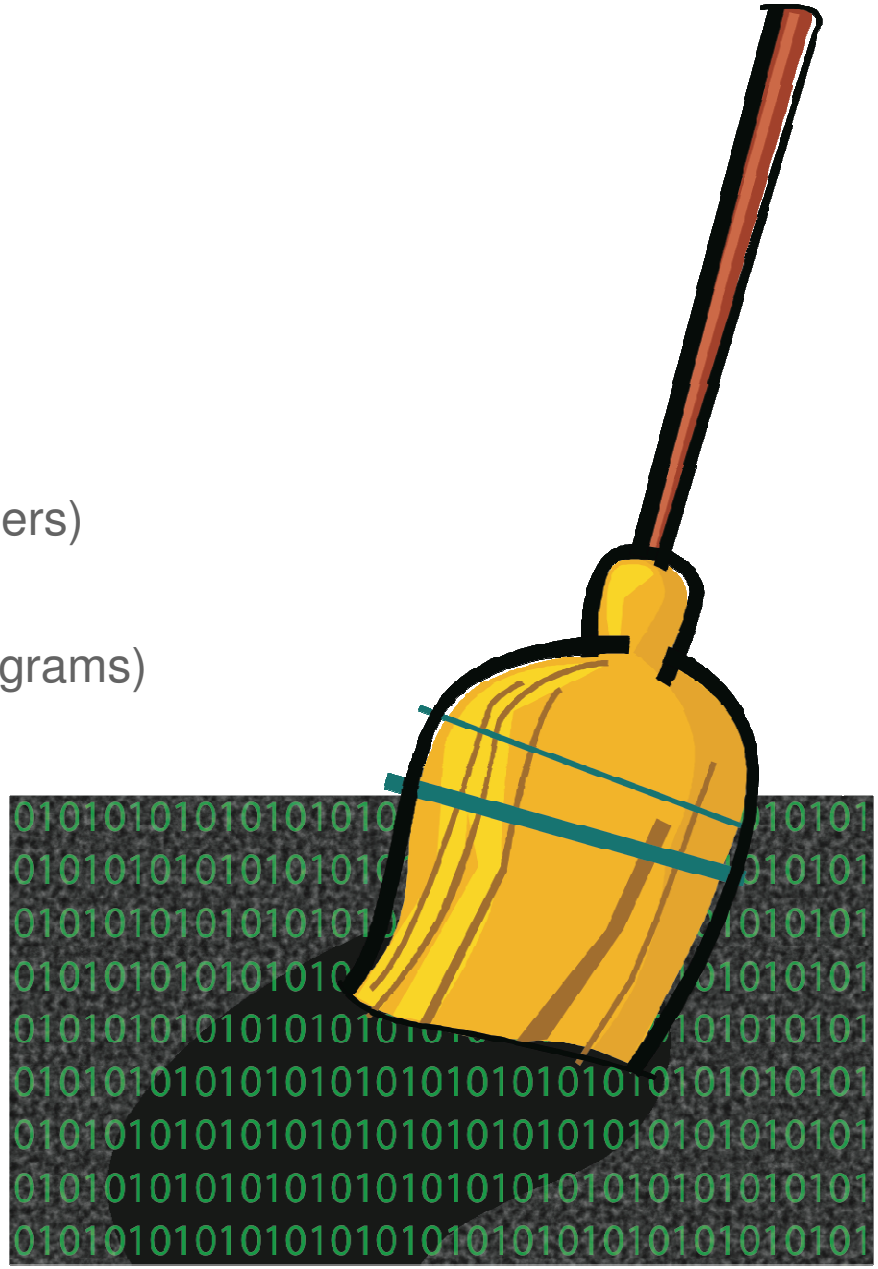


Data Requests

- Can take a long time to get data
- Information Warehouse analysts frequently very busy with a variety of different kinds of requests
- Analytics requests tend to be more complex
- Identifying the right data elements can be time consuming
- Clinical data analytics vs. business & financial analytics are very different

Data cleaning

- Record matching
- De-duplication
- Spellchecking
- Data type constraints (numbers vs. letters)
- Range constraints (plausible values)
- Unit consistency (ounces, pounds, kilograms)



Deleting problematic data

- Removal of duplicates
- Removal of invalid entries

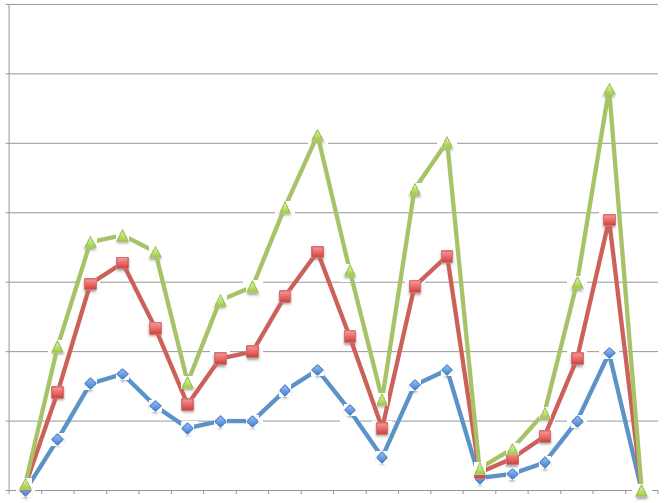
- Questions to consider
  - Are the invalid entries really invalid and not just data in which the semantics are unknown? (e.g., Out of range data, NULL, or empty?)
  - Will deleting such entries result in a systematic bias to the resulting analysis?

Chart Review

- Likely will need to do some manual chart review
  - What data is actually being regularly documented
  - Figuring out what chunks of data are going where
  - Figuring out what chunks of data are missing or difficult to extract

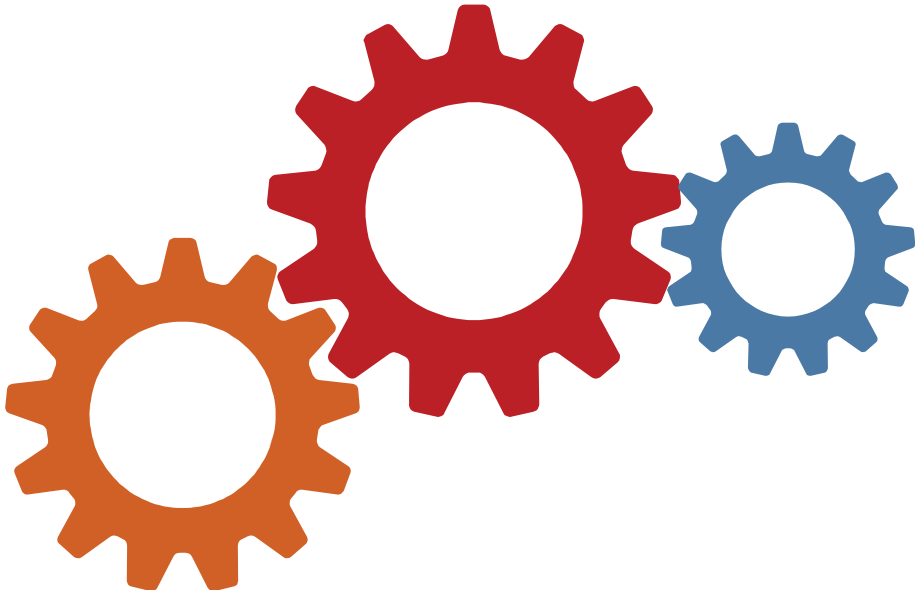
Longitudinal Analysis Challenges

- Things to consider when doing longitudinal analyses
- Things change over time, e.g.
  - Best practices change
  - Medications and procedures change
  - Coding schemes change (e.g., comparing ICD-9 and ICD-10 will be challenging)
  - Comparing data captured when on a previous EMR to your now integrated EMR system will be challenging



Data Integration Challenges

- Integration of data and analyzing data across practice sites and specialties is challenging due to:
  - Practice variation
  - Coding differences
  - Documentation differences
    - structured & unstructured data



Heart Failure Readmissions

- Risk stratification of patients
  - Challenges
    - Extraction of ejection fraction information from notes
    - Social and family history

Things to consider

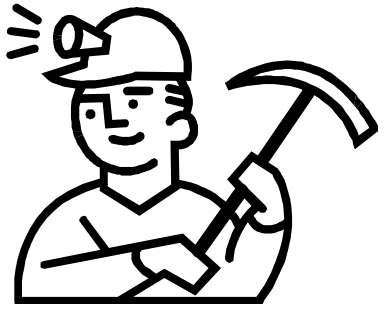
- Dates and times
  - Many different dates to choose from, e.g.
    - Admission time
    - Note times
      - Note started
      - Note finalized
    - Lab information –
      - Order date/time
      - When specimen was collected
      - When result was available
      - When reviewed

Working with Diagnoses & Problem Lists

- Primary diagnosis? Or on the list of diagnoses?
- Diagnoses don't always up to date
- Items are frequently not removed from problem lists
  - e.g., a specialist may add something to a problem list, the problem may be resolved, but then the next physician to look at the record may not remove it from the problem list

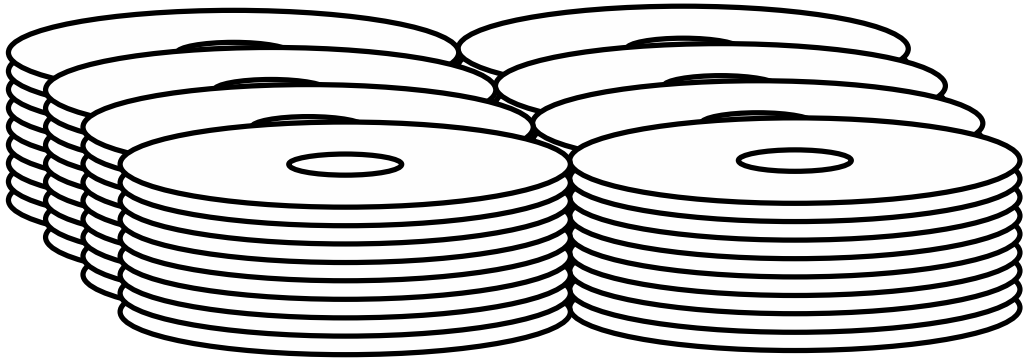
Computational Techniques Needed to Support Data Analytics

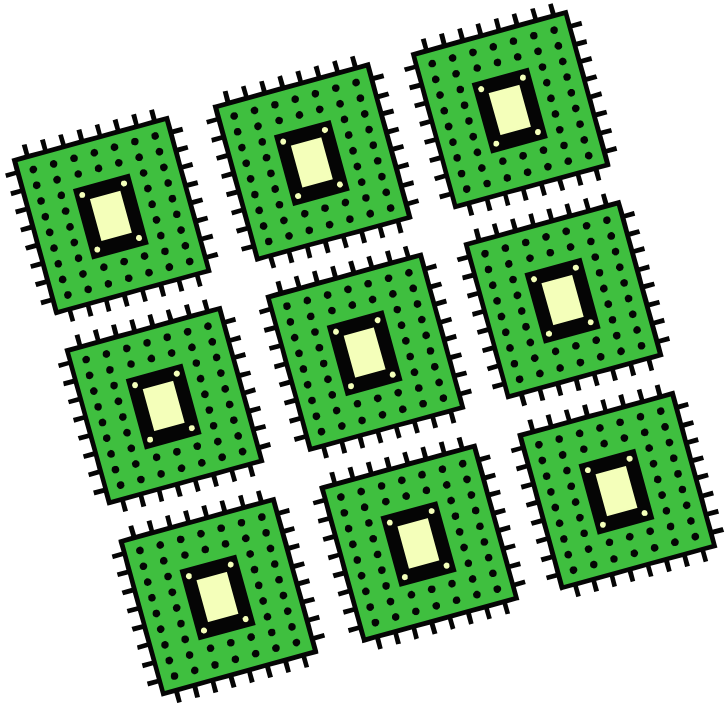
- Natural Language Processing
  - Many NLP engines designed & trained on Newswire text
- Statistical analysis
  - Challenge: data sources not as clean as traditional research data sources
- Machine learning & Data Mining
  - Techniques used in other domains don't always work & need to be modified



Computational Resources For Advanced Analytics Techniques

- Increased disk & CPU
- Increased database resources
- Large compute needed to process and index text

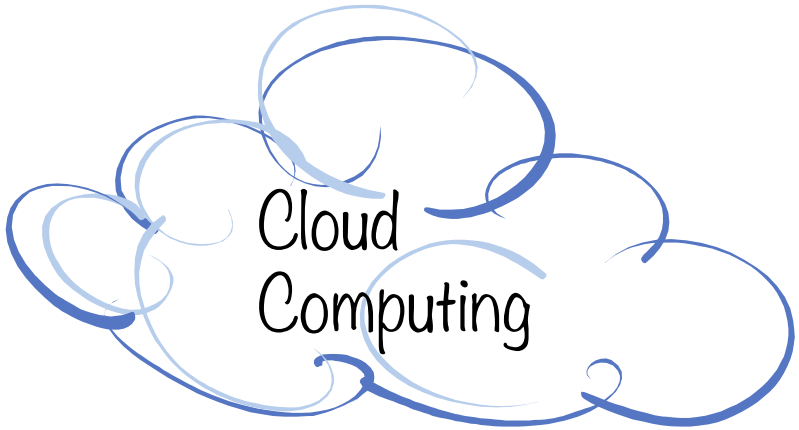




Cloud and Analytics

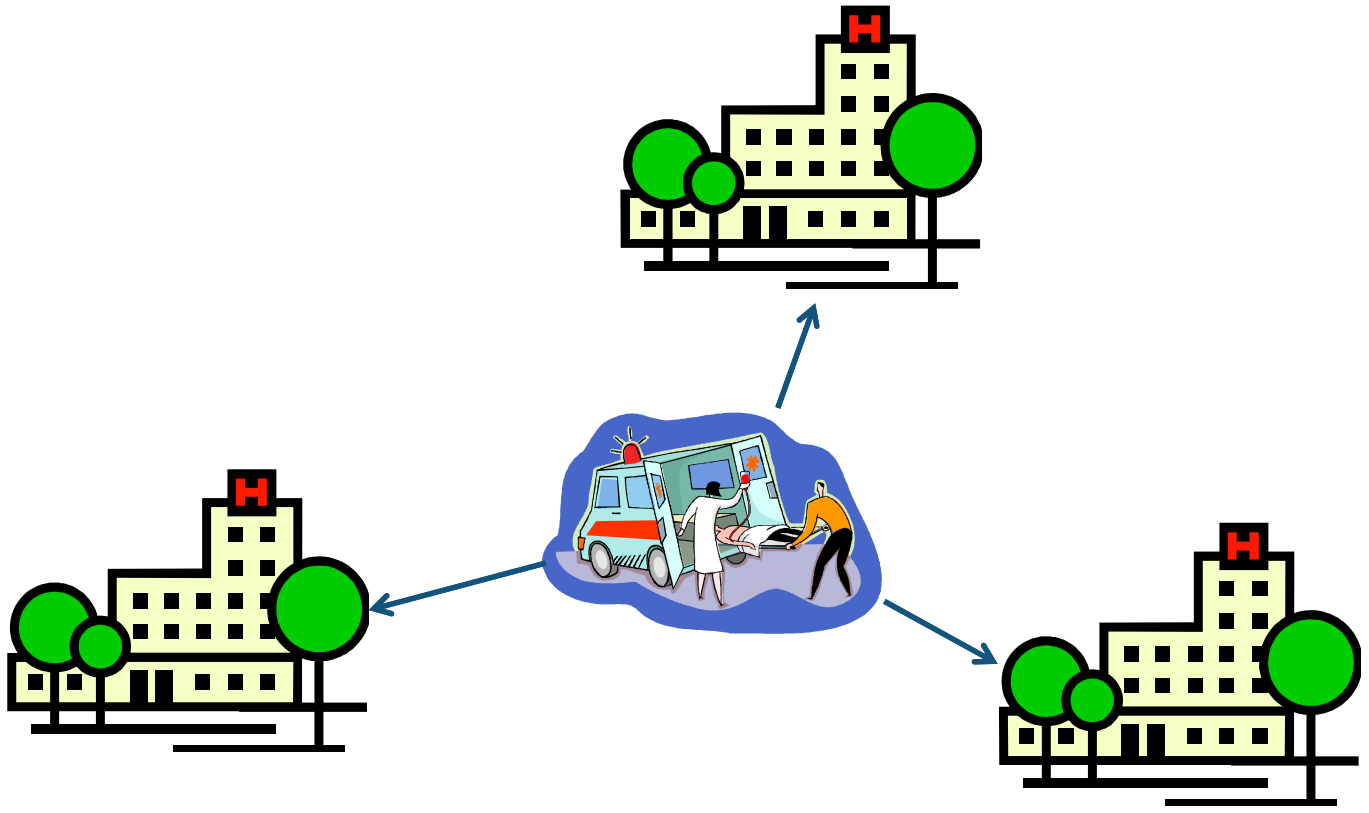
- Computing in the cloud gives access to large computational resources on demand.

- Challenges with using cloud computing and data analytics
  - Moving data to the cloud (can take a lot of bandwidth & time)
  - Leaving data in the cloud has a lot of cost associated with it
  - Establishing HIPAA BAAs



Challenges with analyzing cost

- Hard to find total cost, especially if a patient is readmitted elsewhere



Moving into Predictive Analytics

HIMSS
CENTRAL & SOUTHERN OHIO *Chapter*

Real-time health care analytics & decision support

- Currently most EHR systems are unable to do advanced analytics inside their engines
- Currently need to execute analytics outside of the EHR and bring the information back for decision support

Challenges for Predictive Analytics

- Home-growing a model can be difficult
  - The ‘n’ at a given institution usually isn’t big enough for a robust predictive model
  - Need to analyze data across large HIEs or a state
- Buying analytics models from a vendor
  - Frequently not validated on a clinical sample
  - Is it going to be applicable at your hospital?

Data not in the traditional record

- Much of the data we want in order to do predictive analytics may not currently be in the record or is after discharge
  - Heart failure
    - Daily weight
    - Daily BP
    - Heart rate
    - Diet
    - Exercise
    - Medication adherence

A Look Towards the Future

- Genomics and Personalized Medicine
- Increased need for customized flowsheets for specialties
- More disk space and computational resources
- More ETL processes into data warehouses for easier data analytics processing
- Mining data across HIEs

How might analytics change IT and IT organizations in hospitals

- Need for larger and larger computational & storage capacity
- Move from virtualized infrastructure to big data infrastructure such as clusters and high performance computing architectures
- Move from transactional data structures to more data warehousing-type data structures
- Need for data scientists, not just DBAs and analysts

Success Stories

- IBM Watson – demonstrated to be accurate in utilization management 90% of the time (enables health insurers to determine which treatments are fair, appropriate and efficient and, in turn, what it will cover)
- Kaiser Permanente examined the frequency of blood clots among women who were prescribed oral contraceptive. They found one formula containing drospirenone increased the likelihood of blood clots by 77% compared with other oral contraceptive formulas.
- Seton Healthcare Family - Identifies high risk CHF patients likely for re-admission by using NLP to extract key elements from un-structured H&P, discharge summaries, echo reports, and consult notes. Identified lack of emotional support and a bulging jugular vein as high predictors.

Questions?