Health Care Data

Health Care Information Systems: A Practical Approach for Health Care Management
Karen A. Wager I Frances Wickham Lee I John P. Glaser

Hye-Chung Kum (kum@tamu.edu)
Associate Professor
Population Informatics Lab (https://pinformatics.org/)
Course URL: http://pinformatics.org/phpm631

PUBLIC HEALTH TEXAS A&M UNIVERSITY | POPULATION INFORMATICS

Learning Objectives

POPULATION INFORMATICS

- Health Care Data and Information Defined
  - What are health data and health information?
- Health Care Data and Information Sources
  - Where does health data originate and why?
  - When does health care data become health care information?
- Health Care Data Uses
  - How do health care organizations use data?
  - What is the impact of the trend toward analytics and big data on health care data?
- Health Care Data Quality
  - How does the quality of health data impact its use?

Johns Hopkins University Reimagines the M.B.A.

… threading health-care courses throughout its program, refocusing on quant skills

- WSJ 1/22/2020
- https://www.wsj.com/articles/johns-hopkins-university-reimagines-the-m-b-a-11579689000
- Part of the research process Mr. Ferrari launched involved asking intensive questions of companies about what specific skills they were looking for when they hire newly minted M.B.A.s. A key finding from the research:
- Employers said they needed M.B.A.s with data-science and data-analytics skills.
- While students don't need to be data scientists, they need to be able to supervise a team of them, Mr. Triantis said. That means they have to understand what certain software and tools are capable of, what analyses can be done with them and then make solid decisions based on sifting through data.
- More broadly, employers want to see an emphasis on science, technology and math skills in combination with softer skills like leadership that have been so in vogue in recent years.

6

6

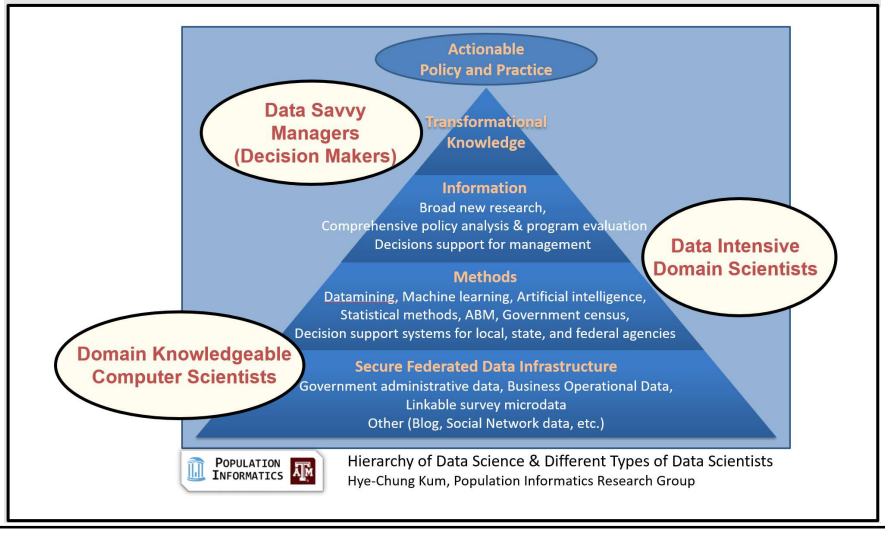
Definitions

- Health Care Data v. Health Care Information
  - Data need to be processed to be information
  - On a continuum, not clear cut
    - One person's data maybe be another person's information depending on the purpose
  - Protected Health Information (PHI)
    - collected by covered entity
    - For covered functions?
- Electronic Medical Records (EMR) v. Electronic Health Records (EHR)
- Scope of Care
  - Episode of Care: a specific condition for a specific time period
  - Continuum of Care: a system that guides and tracks patients over time through a comprehensive array of health services spanning all levels and intensity of care
  - Population Health: varies. Managing population health is to improve health outcomes within defined communities

7

7

Data to Decision



8

8

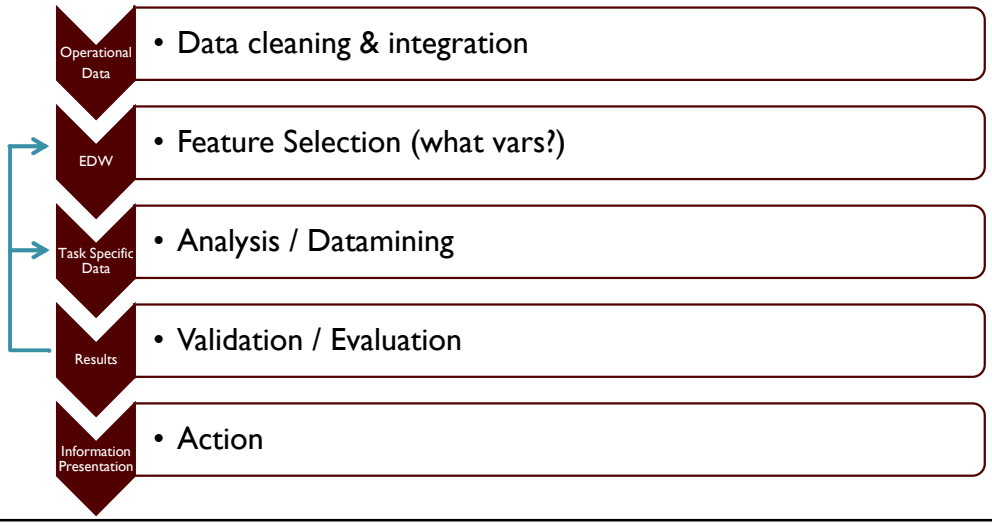
Data Science
Knowledge Discovery & Data mining (KDD)

Big Data : operational data

KDD: Clean, Merge, Reprocess

9

KDD Process



10

Purposes: Patient Records

- Patient care
- Communication
- Legal documentation
- Billing and reimbursement
- Research and quality management
- Population health
- Public health

11

11

Components: Patient Records

- Identification screen
- Problem list
- Medication record
- History and physical
- Progress notes (SOAP)
  - Subjective findings, Objective Findings, Assessments, Plan
- Consultation: outside provider
- Physician's orders
- Imaging and x-ray reports
- Laboratory reports
- Consent & authorization forms
- Operative reports
- Pathology reports
- Discharge summary

12

12

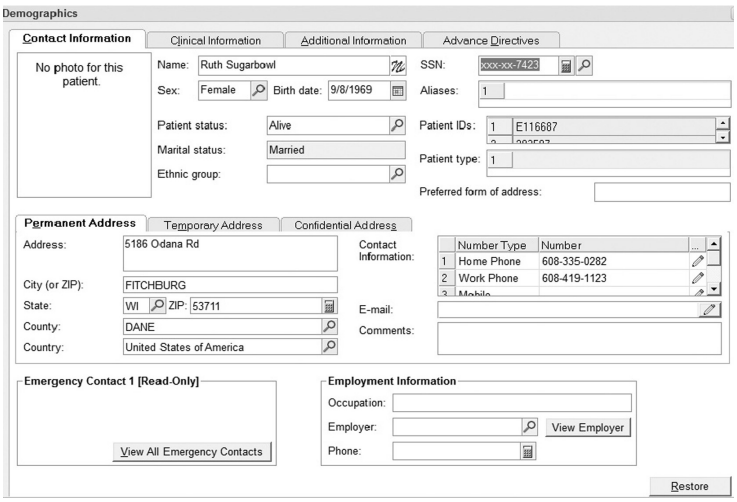
Components: Claims

- Accounting/Billing
  - Verify insurance coverage
  - Bill third party payers
    - UB-04/CMS-1450 (837I): institutional (hospitals, SNFs, home health, FQHC, community mental health centers etc)
    - CMS-1500 (837P): non-institutional provider claims (private physician services)
      - ✓ AMA
  - Process payments upon receipt

13

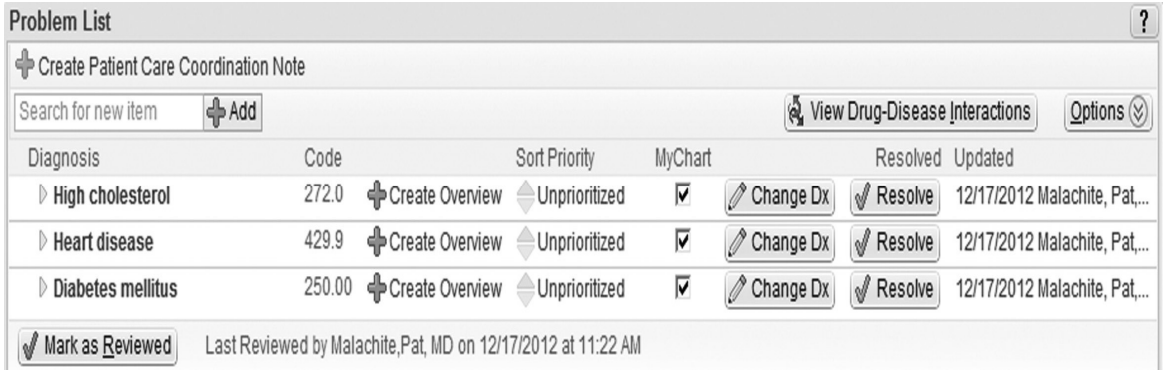
13

EHR Information Screen



Source: Medical University of South Carolina; Epic.

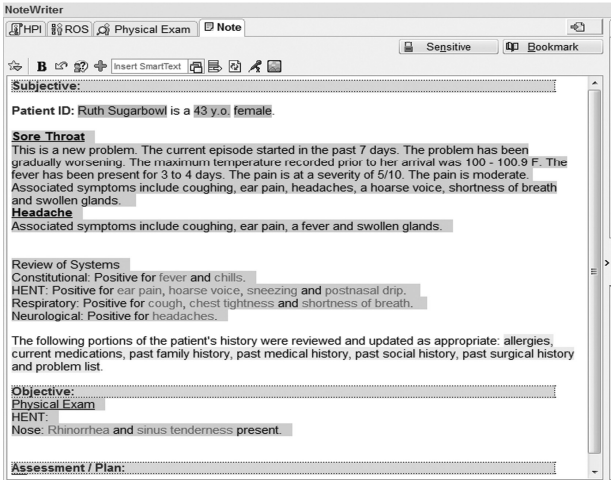
EHR Problem List



Source: Medical University of South Carolina; Epic.

EHR Progress Notes

POPULATION INFORMATICS



Source: Medical University of South Carolina; Epic.

16

16

EHR Lab Report

POPULATION INFORMATICS

Results — CBC and differential (Order 23836)

Procedure: CBC and differential | Abnormality | Status

CBC and differential — Status: **Final result** MyChart: **Not Released** Next appt with me: **None** Dx: **Diabetes Mellitus**

	Range	3yr	3yr
HEMOGLOBIN	13.5 - 17.5 g/dL	13.0 (A)	12.0 (A)
HEMATOCRIT	41 - 53 %	36 (A)	44
RED BLOOD CELL COUNT	4.50 - 5.90 10^6/µL	4.60	4.89
MEAN CORPUSCULAR VOLUME	82.0 - 108.0 fL	89.0	91.2
WHITE BLOOD CELL COUNT	10^3/mL	8.4	7.4
LYMPHOCYTES MANUAL		2	2
EOSINOPHILS MANUAL		0	0
PLATELET COUNT	150 - 399 K/µL	350	389
MEAN PLATELET VOLUME	7.5 - 11.5 fL	9.8	9.9

Resulting Agency
Lab Flowsheet
Order Details | View Encounter | Lab and Collection Details | Routing | Result History
Last Resulted: 10/22/09 3:38 PM

Source: Medical University of South Carolina; Epic.

17

17

Codes: Diagnostic and Procedural

- ICD-10 (International Classification of Diseases)
  - ICD-10-CM (clinical modification) vs ICD-10-PCS (procedure coding system)
- CPT (Current Procedural Terminology)
  - Copyrighted with all rights to publication and distribution held by the AMA
  - Provides a uniform language for describing medical and surgical services
  - HCPCS level 1: standard for physician's office, outpatient, ambulatory care
- Coding plays a major role in reimbursement for care
  - Both up coding and down coding is a problem

18

18

Example ICD10 ?

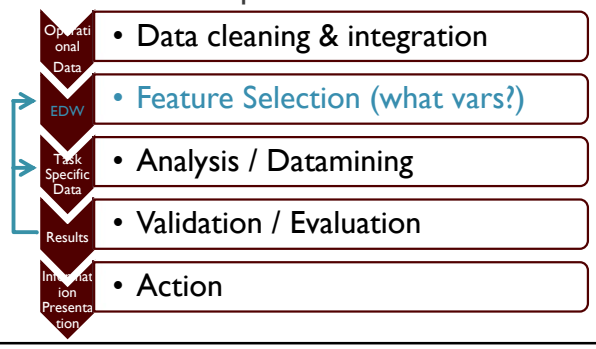
- Our daughter is a coder in Indianapolis and with the news about the iguanas falling out of trees in Florida with the cold spell there, I asked her
- Question: Is there was an ICD10 code for someone showing up in an ER after having been hit by a falling iguana.
  - Answer Y/N ?
- Question: If so, what is it?
  - Answer ?
- There is: Struck by other non-venomous reptiles code W5982XA.

19

19

Data Warehousing considerations

- What data to include?
- How to reconcile inconsistencies?
- How often to update?



20

Accountable Care Act

- Shifting focus from episodic care to population health
- Successful population health require extensive coordination of care across providers and community organizations
  - Comprehensive shared care plan (CSCP)
  - Care managers are needed to interact with patients on a regular basis during and in between clinical encounters
- Reliance on HIT

21

21

Healthcare Data Quality

Depends on the use of the data

- Traditionally (file cabinet):
  - Patient Clinical/Claim Records
    - Episodic
    - Generally from a single organization
- Today (conveyor belt - big data):
  - EHR/Electronic Claims Record
    - Continuous instead of episodic
  - Criteria for quality has shifted

22

22

Small v. Big Data

- Small Data
  - Comparable to a filing cabinet
  - Static
- Big Data
  - Comparable to a conveyor belt
  - Volume
  - Variety
  - Velocity
  - Veracity

23

23

Properties of BIG DATA : 4V

- Volume : constantly generating
- Velocity : constantly changing
- Variety : expressed in many ways
- Veracity : lots of errors
- (Value)

EXAMPLE: the INTERNET!
What do you do to find information/knowledge on the Internet?

24

Small v. Big Data Examples

- Small Data
  - Disease and Procedure Indexes
  - Health Care Statistics
    - Utilization Statistics
    - Performance Statistics
- Big Data
  - Collected from a variety of sources
  - Wide range of uses across multiple industries and efforts

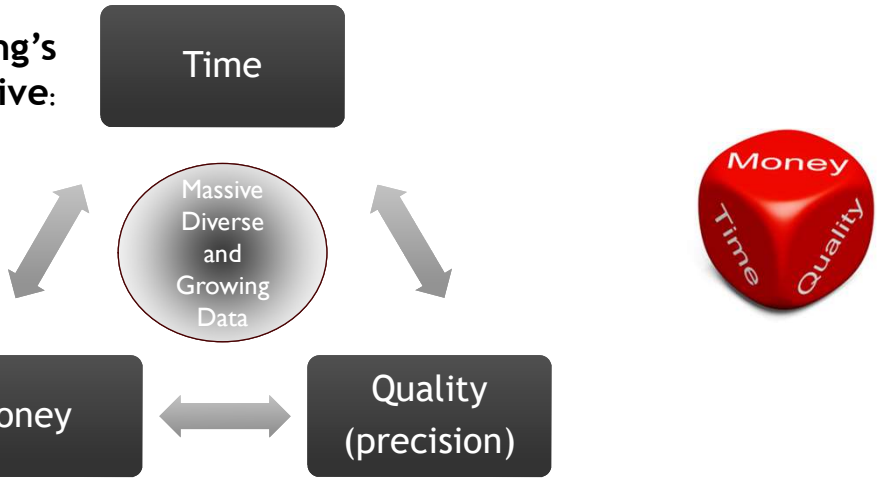
25

25

The Big Data Problem – Nutshelled
Michael Franklin (UC Berkley)

POPULATION INFORMATICS

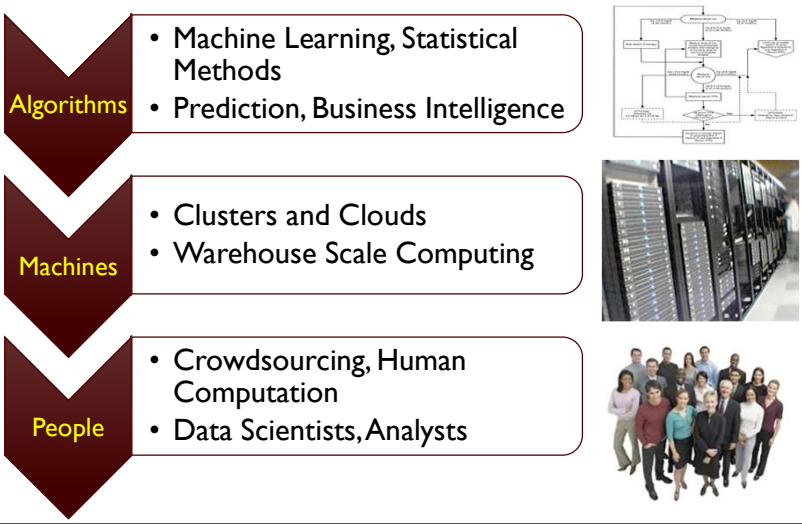
Something's gotta give:



26

AMPLab: Integrating Three Key Resources

POPULATION INFORMATICS



27

AHIMA Data Quality Characteristics

- Accuracy
- Accessibility
- Comprehensive
  - all relevant without inundating
- Consistency
- Currency
  - Data go obsolete
- Clear Definition
  - Used and understood to be the same
- Granularity: unit of data
- Precision
- Relevancy
- Timeliness

Q: Accuracy vs Precision ?

28

Five dimensions of EHR data quality Weiskopf and Weng

- Fundamental
  - Completeness
  - Correctness: free of error
  - Currency
- Proxies to assess fundamental dimenions
  - Concordance: consistency, reliability
  - Plausibility: believeability, validity

29

Strategies for minimizing data quality issues 'Garbage in Garbage out'

POPULATION INFORMATICS

- Up front vs after the fact
- Standardizing data entry fields
- Designing data elements to avoid errors
  - Checking digits, algorithms, well-designed user interfaces
- Developing and adhering to guidelines for documenting the care that was provided
- Building human capacity, including training, awareness building

30

30

Read Assignment 2 & Lab 2

31

Video: Reported in a usable manner

- TX Mental Health Landscape (2:46)
  - https://www.youtube.com/watch?v=8dPqQt0yXJA
- Wealth Inequality (1:30)
  - https://www.youtube.com/watch?v=QPKKQnijnsM

32

Data Analysis: Assignment 1

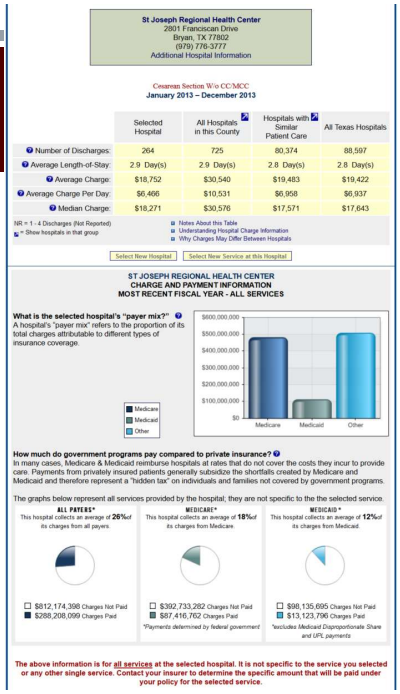
Four basic elements of data analysis

- Source of data
  - EHR, claims data, laboratory data, etc.
- Stored in a retrievable manner
  - Database or data warehouse
- Analytical tool applied
  - Mathematical statistics, probability models, predictive models, etc.
- Reported in a usable manner

33

DSS (Reports)

- PricePoint
  - for consumers
- http://www.txpricepoint.org/
- http://www.txpricepoint.org/Report.aspx?DRG=766&FacilityID=0410500



St Joseph Regional Health Center
2801 Franciscan Drive
Bryan, TX 77802
(979) 776-3777
Additional Hospital Information

Cesarean Section W/o CC/MCC
January 2013 – December 2013

	Selected Hospital	All Hospitals in this County	Hospitals with Similar Patient Care	All Texas Hospitals
Number of Discharges:	264	725	80,374	88,597
Average Length-of-Stay:	2.9 Day(s)	2.9 Day(s)	2.8 Day(s)	2.8 Day(s)
Average Charge:	$18,752	$30,540	$19,483	$19,422
Average Charge Per Day:	$6,466	$10,531	$6,958	$6,937
Median Charge:	$18,271	$30,576	$17,571	$17,643

NR = 1 - 4 Discharges (Not Reported)
* = Show hospitals in that group

Notes About this Table
Understanding Hospital Charge Information
Why Charges May Differ Between Hospitals

Select New Hospital | Select New Service at this Hospital

ST JOSEPH REGIONAL HEALTH CENTER
CHARGE AND PAYMENT INFORMATION
MOST RECENT FISCAL YEAR - ALL SERVICES

What is the selected hospital's "payer mix?"
A hospital's "payer mix" refers to the proportion of its total charges attributable to different types of insurance coverage.

Medicare
Medicaid
Other

How much do government programs pay compared to private insurance?
In many cases, Medicare & Medicaid reimburse hospitals at rates that do not cover the costs they incur to provide care. Payments from privately insured patients generally subsidize the shortfalls created by Medicare and Medicaid and therefore represent a "hidden tax" on individuals and families not covered by government programs.

The graphs below represent all services provided by the hospital; they are not specific to the the selected service.

ALL PAYERS*
This hospital collects an average of 26%¢ its charges from all payers.
$812,174,398 Charges Not Paid
$288,208,099 Charges Paid

MEDICARE*
This hospital collects an average of 18%¢ its charges from Medicare.
$392,733,282 Charges Not Paid
$87,416,762 Charges Paid
*Payments determined by federal government

MEDICAID*
This hospital collects an average of 12%¢ its charges from Medicaid.
$98,135,695 Charges Not Paid
$13,123,796 Charges Paid
*excludes Medicaid Disproportionate Share and UPL payments

The above information is for all services at the selected hospital. It is not specific to the service you selected or any other single service. Contact your insurer to determine the specific amount that will be paid under your policy for the selected service.

34

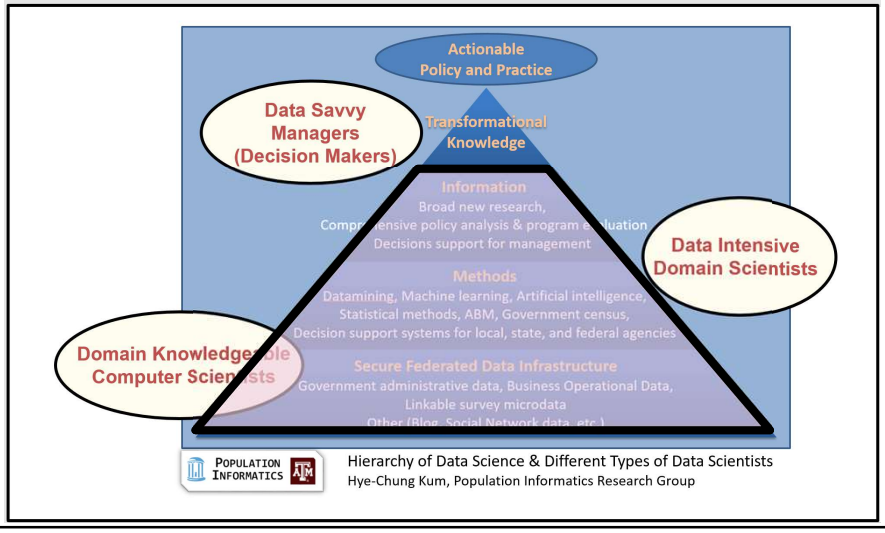
What is Data Mining?

POPULATION INFORMATICS

- Using a combination of artificial intelligence, machine learning, and statistical analysis to analyze data
- and discover useful patterns that are "hidden" there

35

Data to Decision

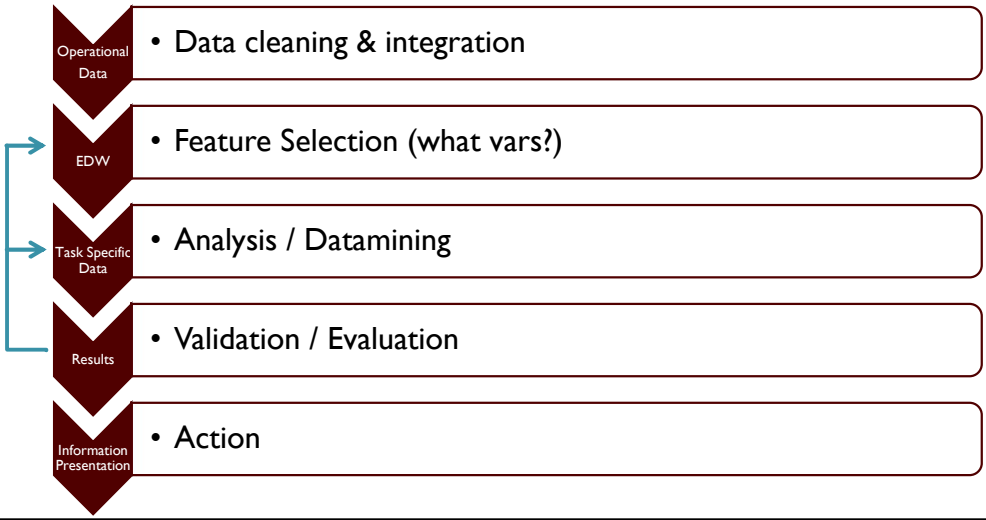


Hierarchy of Data Science & Different Types of Data Scientists
Hye-Chung Kum, Population Informatics Research Group

36

36

KDD Process



37

Sample Applications

- Direct Marketing
  - identify which prospects should be included in a mailing list
  - Clinical trial recruitment: cohort identification
- Market segmentation
  - identify common characteristics of customers who buy same products
  - Profile common characteristics in homogeneous patient group
  - John Billings and Maria C. Raven Dispelling. An Urban Legend: Frequent Emergency Department Users Have Substantial Burden Of Disease. Health Affairs, 32, no.12 (2013):2099-2108
- Customer churn
  - Predict which customers are likely to leave your company for a competitor
  - Potentially Preventable Readmissions to ED
- Market Basket Analysis
  - Identify what products are likely to be bought together
  - Care coordination: common services for a condition (bundled services)
- Insurance Claims Analysis
  - discover patterns of fraudulent transactions (medical fraud)
  - compare current transactions against those patterns

38

Business uses of data mining: Essentially five tasks

- Classification: Group data into predetermined categories
  - Classify credit applicants as low, medium, high risk
  - Classify insurance claims as normal, suspicious
- Estimation: Estimate probability of an event through models built from previous data
  - Estimate the probability of a direct mailing response
  - Estimate the potential cohort size for a clinical trial
- Prediction: Predict an outcome based on input based on models built from previous data
  - Predict which customers will leave within six months
  - Predict which patient will return to the ED
- Affinity Grouping: Group people based on similar characteristics
  - Find out what books to recommend to Amazon.com users
  - Find treatment regime that was successful for similar patient
- Description
  - Help understand large volumes of data by uncovering interesting, useful, and actionable patterns

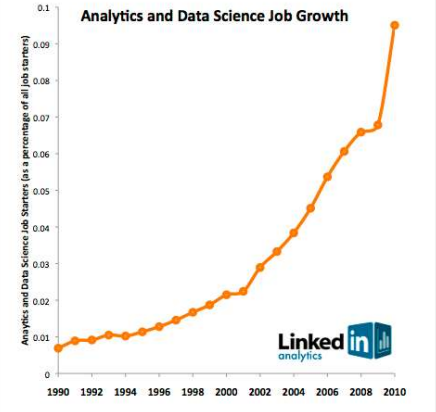
39

Applications in Health

- A March 2014 poll from MeriTalk and EMC found that 63 percent of healthcare executives in the federal government believe that big data will improve population health management
- Examples
  - Manage population health
    - Accountable Care Organizations (ACO)
  - Clinical decision support
  - Cohort identification for clinical trials
  - Medical fraud detection

40

Job market of data scientists



Courtesy LinkedIn Corp.

- statisticians will be the next sexy job
  - Google Chief Economist Hal Varian
- shortage of 190,000 data scientists by the year 2019
  - McKinsey Global Institute

41

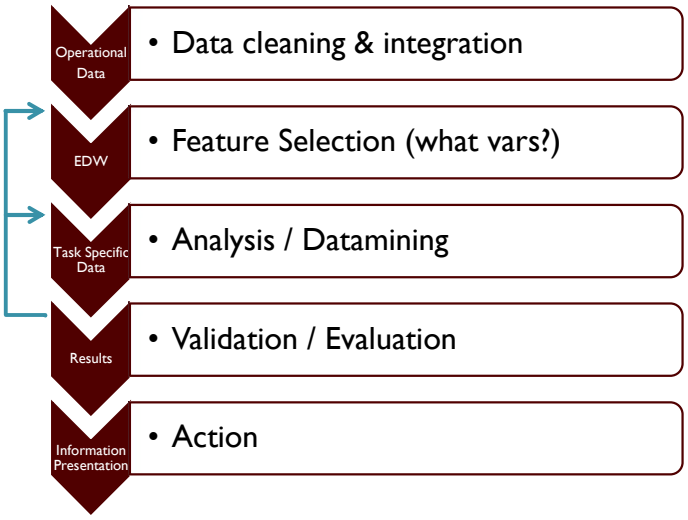
41

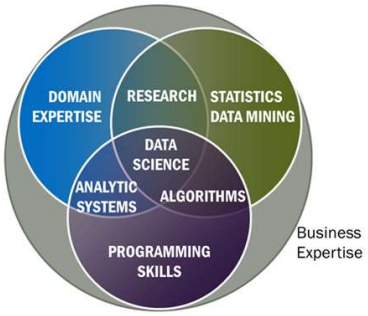
Case Study

- Specialized Program for High Utilizers in One Hospital Network
  - innovative and effective High Alert emergency department overutilization program
  - "SETON's High Alert Program is a specialized, focused case management program. In this program, we develop individualized care plans based on the particular circumstances of individuals. These care plans can be electronically pushed into future clinical encounters to ensure the individualized plan is followed, resulting in consistent care directed at the specific patient."
  - Dr. Christopher Ziebell (Emergency Department Medical Director at the University Medical Center Brackenridge)
  - Dr. Ziebell serves on the Board of Trustees of SETON Healthcare and the Board of Managers of Emergency Service Partners; is Executive Director of Hospital Physicians in Clinical Research; and, chairs the EMS Steering Committee of the Travis County Medical Society.

42

Take Away I
What is Data Science? KDD Process





43

43

Take Away II
What is Big Data ?

- 4 Vs of Big Data
  - Volume : lots of data
  - Velocity : constantly generating & changing
  - Variety : expressed in many ways
  - Veracity : lots of errors
  - (Value)

- Big Data Problems
  - Time
  - Money
  - Quality (Precision)
- Three Resources: AMP
  - Algorithm
  - Machine
  - People

44

44

Take Away III
Business uses of data mining: Essentially five tasks

- Classification
  - Classify credit applicants as low, medium, high risk
  - Classify insurance claims as normal, suspicious
- Estimation
  - Estimate the probability of a direct mailing response
  - Estimate the potential cohort size for a clinical trial
- Prediction
  - Predict which customers will leave within six months
  - Predict which patient will return to the ED
- Affinity Grouping
  - Find out what books to recommend to Amazon.com users
  - Find treatment regime that was successful for similar patient
- Description
  - Help understand large volumes of data by uncovering interesting, useful, and actionable patterns

45

45

How do you get good with data ?

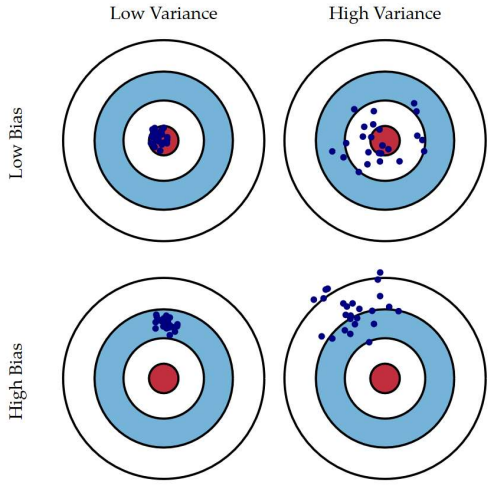
- Sorry, no short cuts. Build experience.
- In this course, start you out.
  - Tableau / Excel
    - Assignment 1
    - lab 1-3, ?++
  - SQL
    - Assignment
    - Labs

46

46

Bias and Variance
http://scott.fortmann-roe.com/docs/BiasVariance.html

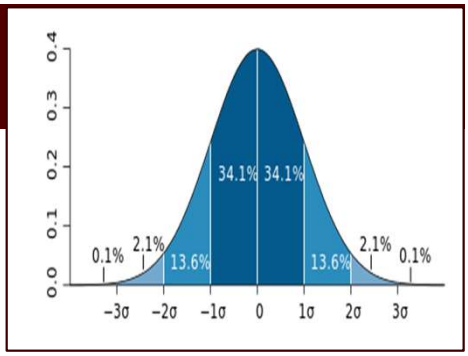
- precise but not valid?
- What is real data like?
- Adjust for bias
- Take into account variance

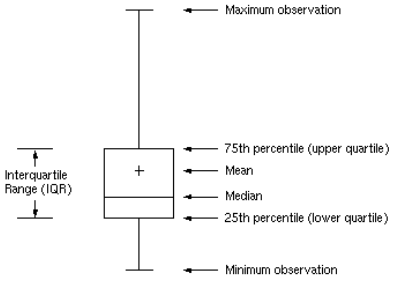


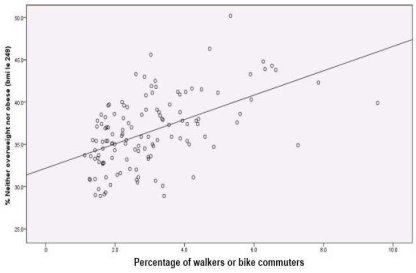
47

Numerical Data : distribution

- Mean
- Standard Deviation
  - How dispersed
- Range: Max/Min
- Median (percentile)
- Scatter Plot: 2 vars



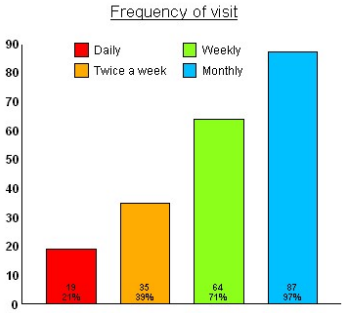


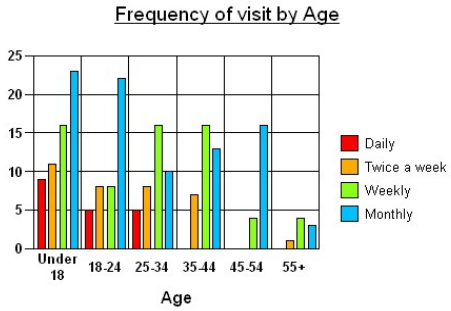


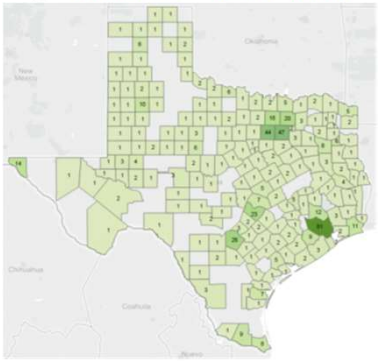
48

Categorical Data

- Tabulation
- Cross tabulation
  - 2 variables
- GIS: maps







49